HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Master's thesis
Geography
Human geography / European Studies Programme

Spatializing Data to Optimize Pricing and Improve User Experience: A Case Study Analyzing Trends in London, Paris and St. Petersburg

Jennifer Leoza Riley

2015

Supervisor:
Dr. Markku Löytönen

HELSINGIN YLIOPISTO
MATEMAATTIS-LUONNONTIETEELLINEN TIEDEKUNTA
GEOTIETEIDEN JA MAANTIETEEN LAITOS
MAANTIEDE

PL 64 (Gustaf Hällströmin katu 2)
00014 Helsingin yliopisto

HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI

Tiedekunta/Osasto Fakultet/Sektion – Faculty Faculty of Science		Laitos/Institution– Department Department of Geography
Tekijä/Författare – Author Jennifer Leoza Riley		
Työn nimi / Arbetets titel – Title Spatializing Data to Optimize Pricing and Improve User Experience: A Case Study Analyzing Trends in London, Paris and St. Petersburg		
Oppiaine /Läroämne – Subject Geography		
Työn laji/Arbetets art – Level Master's Thesis	Aika/Datum – Month and year December 2015	Sivumäärä/ Sidoantal – Number of pages 50

Tiivistelmä/Referat – Abstract

Technology continues to rapidly change the ways in which we live. The ground transportation industry is one of the most recent to be in the midst of this fast paced change.

In this thesis, I analyze data from [redacted], one of the actors within the ground transportation industry. In a three city (London, Paris and St. Petersburg) case study, I spatially analyze the point density distribution of two separate groups of data – booked rides and searched (but not booked) rides.

I hypothesized that each city would show clear differences in the booked versus searched ride data. These differences would spell out specific transfer routes that need to be enhanced in order to increase booking volumes.

According to my results, only one city behaved in the above mentioned fashion. Instead of all three cities behaving the same way, a model of city behavior emerged with three distinct city types.

At one end of the scale is city type number one, where the booked ride patterns closely match the searched ride patterns. There are no specific calls to action here, instead city type number one is the goal for all cities within the [redacted] network. With little mismatch between the searched and booked ride data, there are not factors preventing or discouraging the customers from executing searches into real bookings. St. Petersburg exhibited the characteristics of city type one.

In the middle of the scale is city type number two, where the booked ride patters differ in a number of fundamental ways from the searched ride patterns. In this city type, a number of specific mismatch locations can be identified for further investigation and pricing enhancements. The majority of these enhancements should be achievable through the current supply base for the given city. Paris exhibited the characteristics of city type two.

Finally, at the far end of the scale is city type number three, where the booked ride patters significantly differ from the searched ride patterns. In this city type, there are many broad area mismatch locations. To enhance the situation here, it is necessary to enroll more suppliers to achieve the pricing and availability that is required. London exhibited the characteristics of city type three.

Overall, a model of the city types within the [redacted] network was created and key recommendations were made to help decrease the mismatch exhibited in Paris and London. The next steps from here include the possiblilty of rolling out such an analysis throughout the entire network to understand the specific pattern in all [redacted] cities. To achieve this, focused planning will be required and the data must be more finely processed from the beginning to move towards greater efficiencies. Spatializing data and analyzing geographic trends can be an enormously powerful tool for many businesses looking to gain a competitive edge.

Avainsanat – Nyckelord – Keywords
Density analysis, GIS, spatial mismatch, ground transportation, price optimization

Säilytyspaikka – Förvaringställe – Where deposited
University of Helsinki, Kumpula Library

Muita tietoja – Övriga uppgifter – Additional information

TABLE OF CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

Technology is rapidly changing the way we live. This is not something new, nor is it about to stop anytime in the near future. In recent years, the catch phrase has been *social media* in its various manifestations; Facebook, Twitter, Instagram, etc. Now, the *sharing economy* has caught on and snatched its fair share of the headlines as well. Powerhouses such as Uber and Airbnb, who are currently valued at $41 billion and $13 billion, respectively, (Ramadan et al. 2014; Macmillan et al. 2014) are the darling daughters of this phenomenon, where people can "turn their fallow assets into cash machines…enabled by efficient online platforms," (Alden 2014).

These are just a few examples among thousands through which technology impacts our daily lives. Ground transportation, specifically taxi and car services, is one of the most recent industries to be caught up in this whirlwind of technological change. To clarify, when I discuss ground transportation here, I am not referring to the on demand Uber/Lyft/Hailo manifestations, but a transformation of the way we think about traditional licensed car services.

Logistics has become quite sophisticated to the point where we click one button on Amazon and our camera, book, dress, etc. will arrive at our front door in a matter of days. Logistics is already quite advanced, however, leaders in the field have started to advocate for a new wave of logistics known as a *physical internet* that will further eliminate vast amounts of inefficient and unsustainable practices (Montreuil 2011). If our logistics systems for moving goods are already this sophisticated, then why are the methods of moving people still so far behind?

The transportation of people at the click of a button – while maintaining choice and variety – is the goal and this is just starting to emerge. When you go on a trip to Berlin, odds are you know where to search for flights and hotels (Kayak, Expedia, Booking.com and Hotels.com as some of the industry's leading examples), but where do you find your taxi from the airport to your hotel? Ground transportation is still a fragmented industry, but this is rapidly changing. There are many companies working to eliminate this fragmentation – [redacted] – just to name a few, but so far, only one actor, [redacted], has been able to partner with leaders in the travel industry ([redacted]) to deliver this one-click travel option for a full door-to-door experience (ground transportation, flights and hotels booked together). Additionally, with Dublin based [redacted] (a company specializing in car rental) purchase of [redacted] in April 2015 (Raeste 2015, Cord 2015), the opportunities for [redacted] to deliver ground transportation offerings to a wider audience have soared, specifically through [redacted] existing relationships with over 80 airlines around the globe.

[redacted] is a technology and service provider who works with local taxi companies (suppliers) around the world to deliver a unified and uniform service. In this way, the fragmented ground transportation market is disappearing. Any user from anywhere in the world can go to the [redacted] website and pre-book a taxi in Moscow to pick them up at the airport and bring them to their hotel. There are no concerns about local currency or language, as the ride has already been paid for ahead of time, including tips, and the driver has already been delivered the passenger's destination information.

The ground transportation industry is changing quickly and I have been offered the unique opportunity to perform a geographic analysis on a portion of this emerging data. Many new companies like [redacted] have lots of data that lays fallow and is not analyzed in any systematic way. My goal is to illustrate the power of spatialized data and show how these techniques can help companies make informed decisions for growth and success.

In this thesis, I spatialize booked and searched ride data in three of [redacted] most popular destinations. After spatializing the data, I employ various GIS methods to answer the following questions:

> Based on booked and searched ride results, how can pricing be optimized to encourage continued positive patterns of booked rides? What are the similarities/differences between booked and searched ride clustering patterns? What actions should be performed to convert more searched rides into booked rides?

My hypothesis is that each city has hotspots that customers are interested in booking but current pricing inhibits these bookings from being made.

2. THEORY & BACKGROUND

2.1 Hotspot Analysis

2.1.1 Definition

In the literature, leading scholars in the field have more or less agreed on a common understanding of what a hotspot is. Harries defines a hotspot as, "a condition indicating some form of clustering in a spatial distribution," (Harries 1999: 112). Spencer Chainey, the Director of Geographical Information Science at University College London defines a hotspot as "a geographic area of higher than average crime [*read* phenomenon] or disorder," (Chainey 2012) This same definition has also been adopted by the U.S. Department of Justice (Eck et al. 2005) in their authoritative report on mapping crime.

Wing and Tynon (2006) have additionally indicated three criteria for an area to be considered a hotspot. They write "Others have defined hotspots more specifically but three hotspot criteria generally are acknowledged: frequency, geography and time, (Wing and Tynon 2006).

2.1.2 Hotspot-related Concepts

Hotspot analysis and clustering are closely linked terms and are often used interchangeably within the literature. Grubesic (2010) explains, "Broadly defined, 'clustering' refers to a group of people or things relatively close to each other in geographic space". Eckley & Curtin 2013 indicate that "a spatial cluster…is a geographic point pattern that demonstrates an excess number of events relative to the expected number of events." Furthermore, as seen in Harries' definition and with other scholars, the term 'clustering' is used to define what a hotspot is.

For clarity purposes, within this thesis, 'hotspot' is the preferred and most often used term to describe the defined higher than average geographic areas.

2.1.3 Applications

Location analysis is a broad topic with applications in a wide variety of disciplines reflecting the multidisciplinary nature of geography and GIS in particular (Blaschke & Merschdorf 2014, Hall et al. 2015, Kasimov et al. 2013). In addition to being used in a variety of disciplines, applications of location analysis can be found globally.

Van Haaren and Fthenakis (2011) use location analysis to determine the optimal location for a new wind farm in the state of New York. Similarly, GIS is used to evaluate the best site for a new ski resort in the Rocky Mountains of the western United States (Silberman & Rees 2010). Eckert and Shetty (2011) examine accessibility to grocery stores in Toledo, Ohio and determine the best location for a new retail site to be opened. Cheng et al. (2007) determine the best location for constructing a super shopping mall in Hong Kong.

With respect to hotspot analysis, it is most common to see the technique employed by police departments and law enforcement to learn more about current and prevent emerging crime patterns. Though hotspot analysis is quite prevalent within criminology, it is not solely limited to this field as the discussion below illustrates. As with location analysis, applications of hotspot analysis can be found around the world.

It is quite common to find applications of hotspot analysis within criminology. The methods within the criminology perspective are quite advanced and have been studied by a large group of scholars. Chainey et al. (2008) look at four crime types (burglary, street crime, theft from vehicles and theft of vehicles) in the north of London and evaluate the best method to predict future crime based on current hotspot analysis techniques. Crime hotspots in national forests in Oregon and Washington State are systematically identified by Wing and Tynon (2006) and T. Balogun et al (2014) identify robbery, rape, burglary, cultism and other crime hotspots in Benin City, Nigeria.

In marine biology, researchers have used hotspot analysis to determine key populations of loggerhead sea turtles in the Mediterranean Sea for conservation purposes (Cambie et al. 2012). Scientists in Texas have also used hotspot analysis to research anthrax in white-tailed deer to evaluate the role that hematophagous flies play as vectors for the disease (Blackburn et al. 2014).

Applications of hotspot analysis are also commonly found in public health and epidemiological research. Examples can be found in the studies conducted by Curtis et al. (2014) on West Nile Virus in Houston, Texas and Tobin et al. (2012) on sex exchange among drug users in Baltimore, Maryland.

2.1.4 Methodology: performing hotspot mapping and cluster analysis

As Vuori explains in her thesis on mapping accessibility in the Peruvian Amazon, accessibility can be measured using multiple methods, and choosing the correct one largely depends on the data and type of analysis the researcher wishes to carry out (Vuori 2009). Hotspot analysis is quite similar to accessibility in this respect, with a variety of options to choose from. Within the hotspot analysis context, Grubesic et al. (2014) write, "For many analysts and researchers, difficulties emerge when trying to determine the appropriate technique for a specific problem or context." Marine biologists studying hotspots also reflect this sentiment. Scholars in the field have stated, "Although various spatial analysis techniques for determining hotspots have been developed in recent years…establishing the appropriate methods…still remains a challenge," (Cambie et al. 2012). Selecting the best type of hotspot analysis to perform requires an intimate knowledge of the data and aspects to be studied.

2.1.5 Numerical location analysis measures

Within the literature it is common to see one of the following numerical or statistical analysis measures employed alongside one of the GIS-based analysis measures listed below. Pairing statistical and GIS techniques helps to strengthen the research results. Some of the most common statistical analysis measures found within the literature include

Monte Carlo methods

GI*

Local Moran's I

2.1.6 GIS-based location analysis measures

There is an array of possible GIS methods that can be employed in hotspot analysis. While the list below is not comprehensive, some of the most frequently used methods have been identified.

Four of the most common GIS hotspot analysis techniques include spatial ellipses, thematic mapping of geographic boundary areas also known as chloropleth mapping, grid thematic mapping and kernel density estimation (KDE). A fifth method of point data visual analysis has almost completely disappeared with sophisticated GIS software becoming widely available (Chainey et al. 2008 and T. Balogun et al. 2014). Wing and Tynon (2006) expand upon this latter topic when they explain, "point maps….are limited because multiple events may be mapped on top of one another."

In their analysis on the effectiveness of hotspot mapping Chainey et al. (2008) conclude, "KDE is better at predicting future spatial patterns of crime in comparison to the other most common hotspot mapping techniques." In their paper on West Nile Virus mosquitoes in Houston, Curtis et al. (2014) also conclude that using kernel density estimation is "an adequate approach".

2.2 [redacted]

As previously mentioned, [redacted] is a technology and service provider that works with local taxi companies (suppliers) around the world to deliver a unified and uniform service. Regular taxi, executive car and minibus rides can be pre-booked around the globe with clear, transparent pricing available to the consumer in a number of currencies.

When a customer books a ride with [redacted] it is immediately delivered to the supplier for review. The supplier sees all of the booking details and if they are able to perform the transfer they *accept* the booking. After the booking has been accepted, the customer receives immediate confirmation that their ride has been successfully reserved.

2.2.1 Topics of Analysis

In this thesis, two categories of data will be analyzed; including *booked rides* and *searched rides*. Booked rides can be further broken down into *cancelled, declined, no-show* and *failed* rides. Each of these terms is discussed in further detail below with examples to make the categorizations clear.

2.2.2 Booked Rides

Throughout this thesis, when the term *booked ride* is used, this refers to a ride that has been purchased. This means that a passenger, travel agent, personal assistant etc. has booked a ride directly on the [redacted] website or through one of [redacted] various booking channels. The majority of booked rides are *completed* meaning that the transfer has been performed as agreed (e.g. Taxi picks up customer on time at Point A and delivers them to Point B).

However it is important to note that the category *booked ride* includes *cancelled, declined, no-show* and *failed* rides. Additionally, test bookings, which are performed when a new supplier is about to be launched in the service, have been removed from the booked ride data.

A cancelled ride has been booked, however the customer has canceled it at least two hours before the scheduled pick-up. Plans change and people may alter their vacation plans or decide not to attend a conference anymore. In these cases, the customer is

Declined bookings are those that the customer books; however the supplier is unable to perform the job. A majority of rides in this category are declined because the supplier is already at full capacity during the scheduled pick-up time. As expected, the customer is not charged for these transfers.

A no-show is when a customer does not show up for their ride. Any cancellations made less than two hours away from the pick-up time are treated as no-shows. Additionally, if the passenger cannot be found nor do they respond to [redacted] Customer Service phone calls, the driver will be released after the included waiting time has expired and the ride will be treated as a no-show. As the supplier has performed the job as agreed, the customer is charged in full.

Failed rides are those that have not been carried out according to the contract agreed between [redacted] and the supplier. Most frequently, a ride is a failure because the driver has shown up to a pick-up more than 10 minutes late or has not shown up at all. A transfer is also sometimes considered a failure if the wrong vehicle type has been dispatched, if the conduct of the driver was extremely unprofessional or unsafe or if the vehicle was in severely poor condition.

Finally, test bookings are those made by the [redacted] Supply Team before the supplier is officially launched in the [redacted] service. Test bookings are performed so the supplier becomes more familiar with all of the [redacted] communications and processes before they are to receive real bookings from [redacted] customers. These have been excluded from the analysis because they do not reflect actual customer booking patterns.

This thesis analyzes *all* booked rides that have been made in London, Paris, and St. Petersburg since the launch of [redacted] service in [redacted]. The sum of booked rides for all three cities is [redacted] and individually there have been [redacted] booked rides in London, [redacted] booked rides in Paris and [redacted] booked rides in St. Petersburg.

2.2.3 Searched Rides

The main difference between a booked ride and a searched ride is that a searched ride *has not* been booked. Searched rides are just that – they are rides that have been searched and an offer has been made. This means that a customer has gone directly to [redacted] or one of [redacted] partner sites and has gotten a quote for the price of a ride from Point A to Point B. For some reason, they have decided not to book the ride. The user may have been deterred by the price, they may be shopping around, might want to see what car types [redacted] offers, etc.

In this thesis, searched rides are included because I want to examine the fundamental geographic differences between booked rides and purchased rides. My objective is to analyze if there are differences between the two categories. If there are differences, I will try to determine what is causing these differences with the aim of reducing price on potentially high demand transfers in an effort to increase sales.

This thesis analyzes *all* searched rides that have been made in London, Paris and St. Petersburg since the launch of [redacted] service in [redacted] The sum of searched rides for all three cities is [redacted] and individually there have been [redacted] searched rides in London, [redacted] searched rides in Paris and [redacted] searched rides in St. Petersburg (Table 1).

	Booked Rides	**Searched Rides**
London	[redacted]	[redacted]
Paris	[redacted]	[redacted]
St. Petersburg	[redacted]	[redacted]
TOTAL	[redacted]	[redacted]

Table 1. A comprehensive view of booked rides and searched rides for the three cities included in the case study.

3. STUDY AREA

[redacted] service is available around the globe. As of November 2015, the service is offered in 285 destinations throughout 66 countries on 6 continents (Riley 2015) and increases almost weekly (Kiseleva 2015; Kiseleva 2015; Riley 2015; Riley 2015). With such frequent updates in the supply network and by nature of the sheer volume of destinations, the scope of this thesis simply does not allow for a thorough analysis of each and every destination.

In this section, we examine the selection criteria for the three cities in the case study, explore the city extents determined by the available data and briefly explore a profile of each city to understand its specific background and context on a global scale.

3.1 Selection Criteria: London, Paris & St. Petersburg

To contain the work within the capacity of a master's thesis, a three-city case study format was selected. Three case studies allow for a sufficient amount of variety but at the same time does not overwhelm the scope and subject of the work. Based on the three factors of importance, sample size and variation within the supply base, London, Paris and St. Petersburg were selected as the case study cities (Figure 3).

These three cities were first chosen because of their popularity within the [redacted] network. In terms of booked ride volumes, these three destinations are consistently among the top 10 within the [redacted] network in both the short term (quarters 3-4, 2014, Figure 1) and long term (two years; 2013-2014, Figure 2). As the aims of this study are to optimize pricing and improve user experience, it makes sense to perform this analysis on top destinations where the results and recommendations will have the largest potential impact on future sales and growth.

[redacted]

Figure 1. Top 10 booked destinations from [redacted] (Niemi 2015).

[redacted]

Figure 2. Top 10 booked destinations from [redacted] (Niemi 2015).

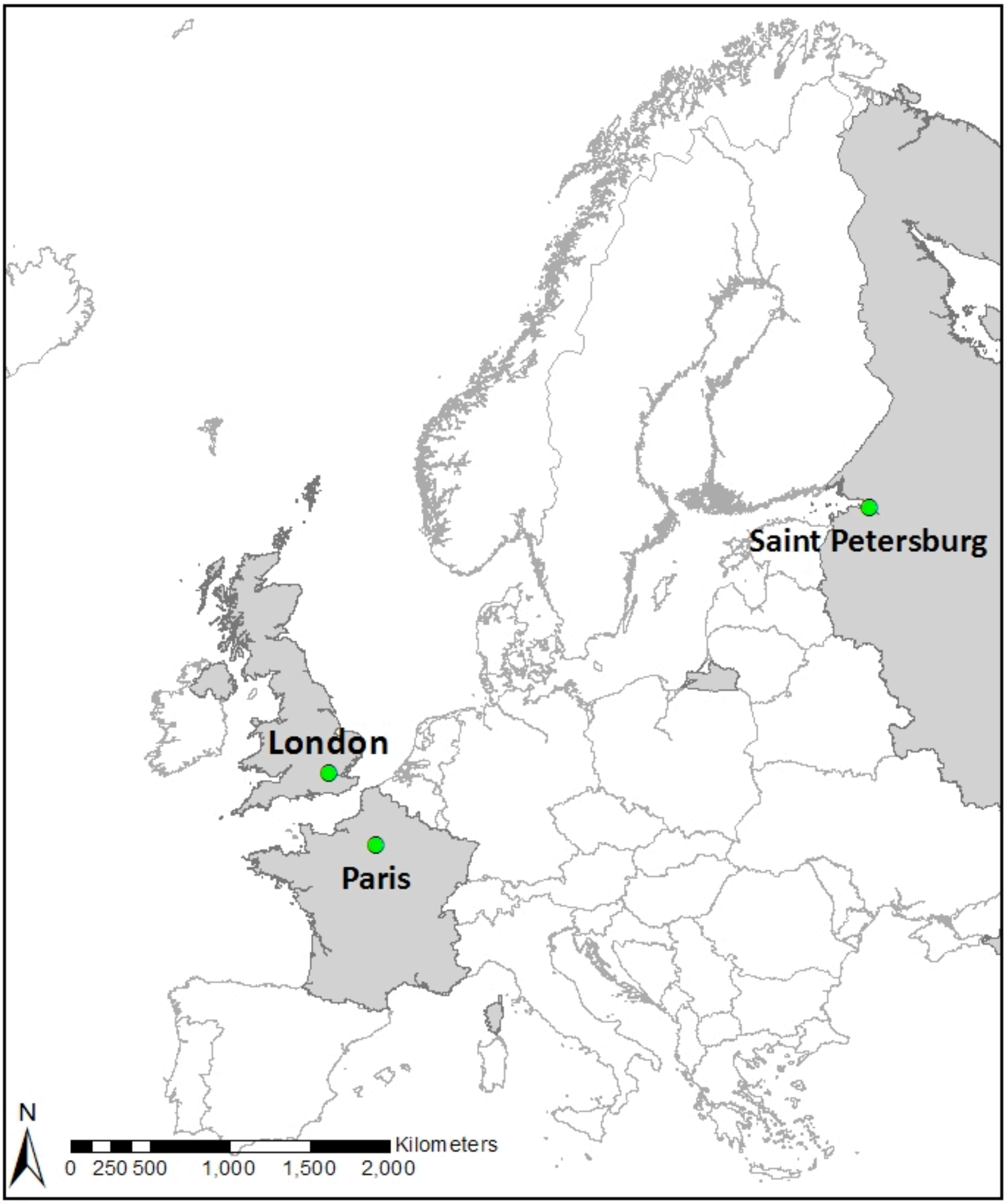

Figure 3. Study area showing the location of the three case study cities (London, Paris and St. Petersburg) within Europe.

Additionally, these three cities were chosen because of their relatively strong sample size within the available data. Using cities at the top of the list allows for a more robust, healthy and diverse set of data. Indeed, 100s of rides have taken place in each city and having a larger sample size will help to flesh out actual booking trends that are taking place within each city. Performing this analysis on cities with a smaller sample size may not show all booking trends or the trends present may, for example, be heavily influenced and swayed by the actions of one client.

Furthermore, the geographic spread of the cities chosen displays a decidedly European centric as opposed to global booking distribution pattern. This is largely a legacy of [redacted] initial business focus, which sought to serve major destinations in Europe (Cutler 2013; Fox 2013). While [redacted] offerings have since significantly expanded well beyond the initial European base (Holmgren 2014, Hansson 2014) the actual booking sales have not yet adjusted to reflect this global distribution. The available data supports an analysis of European destinations as opposed to top global cities.

Finally, London, Paris and St. Petersburg were chosen because of their variations within the supply base. As of March 2015, the majority of [redacted] cities were covered by the services of one supplier. However, in some cities, the coverage comes from a larger network of suppliers deemed necessary for successful price, quality and coverage competition. The three cities in this study represent the full range of supplier activity within the [redacted] network. St. Petersburg has primarily had one supplier, the Parisian supply base is just starting to expand and we have had more suppliers in London than any other city in our network. For a summary of suppliers in each city that this analysis is based on, see Table 2. Having the analysis performed on these selected cities makes the data more representative of the [redacted] network as a whole.

City	Supplier	Cooperation Start Date	Cooperation End Date
London			
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]		
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
Paris			
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]
St. Petersburg			
	[redacted]	[redacted]	[redacted]
	[redacted]	[redacted]	[redacted]

Table 2. A list of suppliers from the representative cities of this case study and their dates of operation.

3.2 City Extents
The city extent is determined by the actual data available for each of the three study cities. Any searched or booked ride associated with suppliers from these three cities have been included. Below, the maximum extent for each city has been identified (Figures 4, 5 & 6). The maps include both the start and end points for both searched and booked rides.

Of all three cities, the extent of the London data was the largest. The core is clearly around the city of London itself, but it extends as far north as Aberdeen and touches upon many areas of Great Britain in all cardinal directions. Determining the specific city area for London was more complex than both Paris and St. Petersburg. In the latter two cities, there are specific suppliers servicing the city and it is quite easy to pull that data for analysis. However, in London, one of the suppliers, [redacted], serves the entire country and picking out the appropriate data to be included for analysis required more robust selection criteria. The selection criteria will be highlighted in further detail in the Materials and Methods section below.

Compared to London, the data extents of Paris and St. Petersburg were much more compact and did not stray far away from the central city core. This can partially be explained as a function of the operational areas of the existing suppliers. This means that the data points can only go so far as the outer extent of the operational areas of the suppliers. Even if the customer would like to book beyond the operational area of the supplier, this would not show up in the data because it was not possible to make the customer an offer.

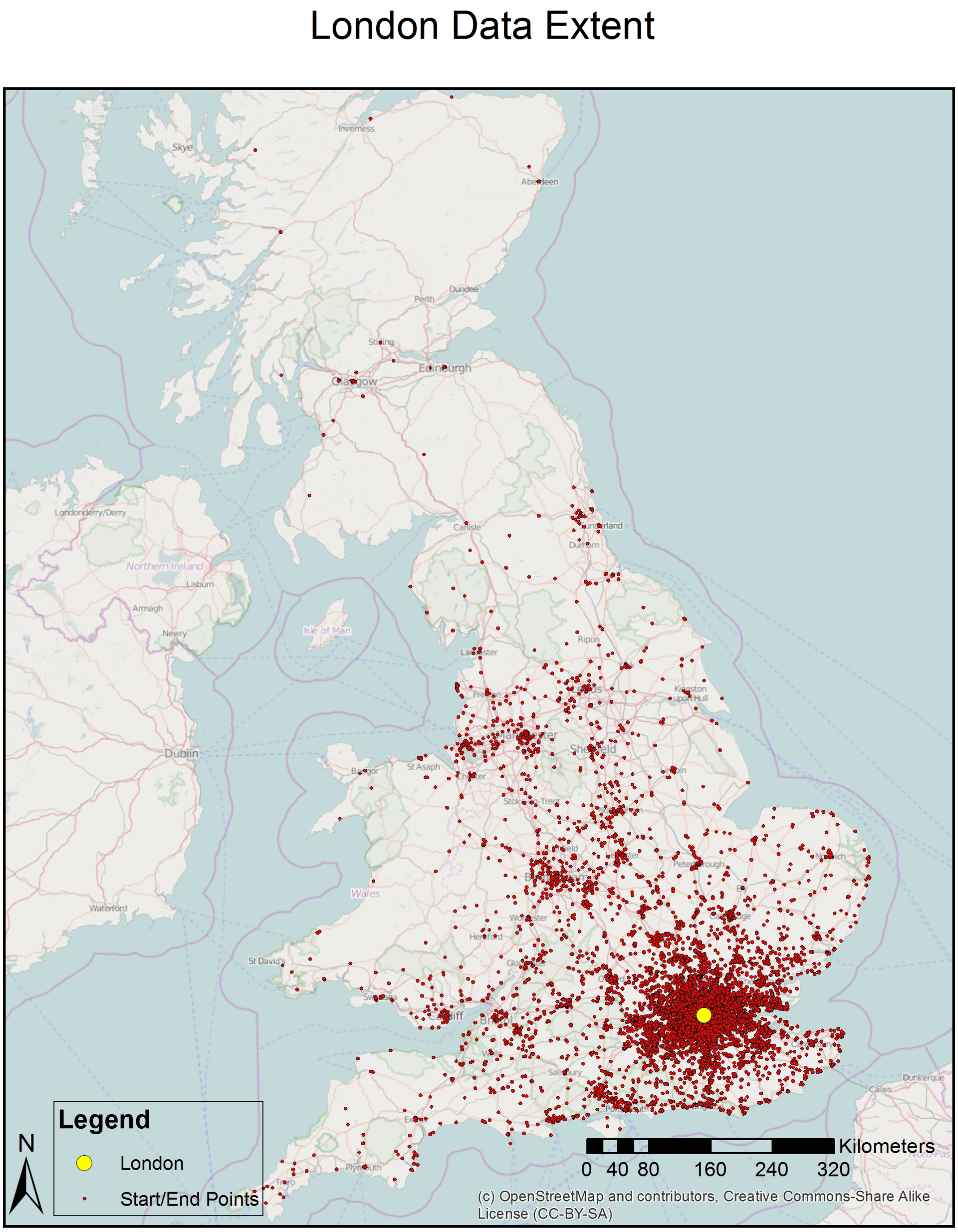

Figure 4. The full extent of the London data including both start and end points for both booked and searched rides.

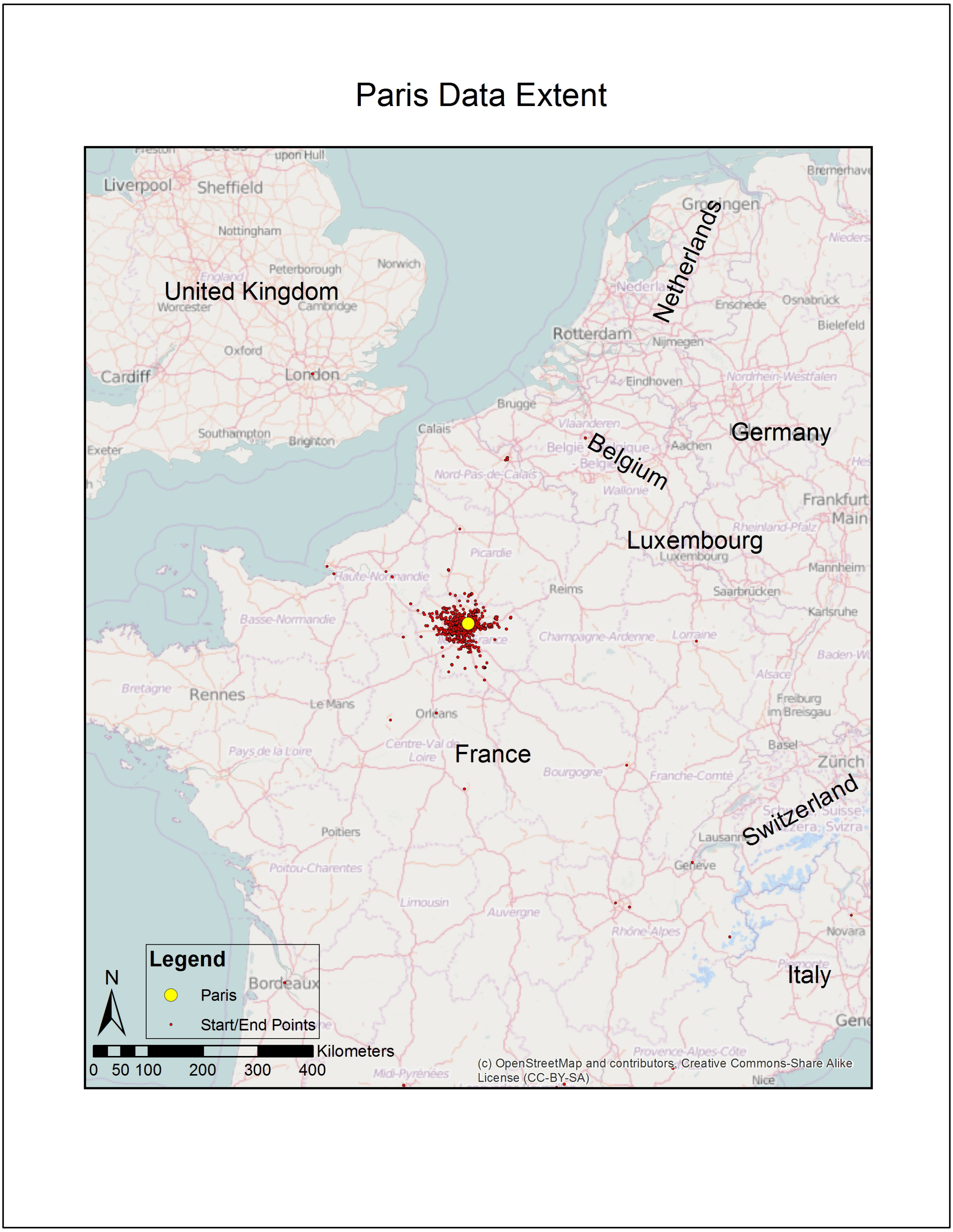

Figure 5. The full extent of the Paris data including both start and end points for both booked and searched rides.

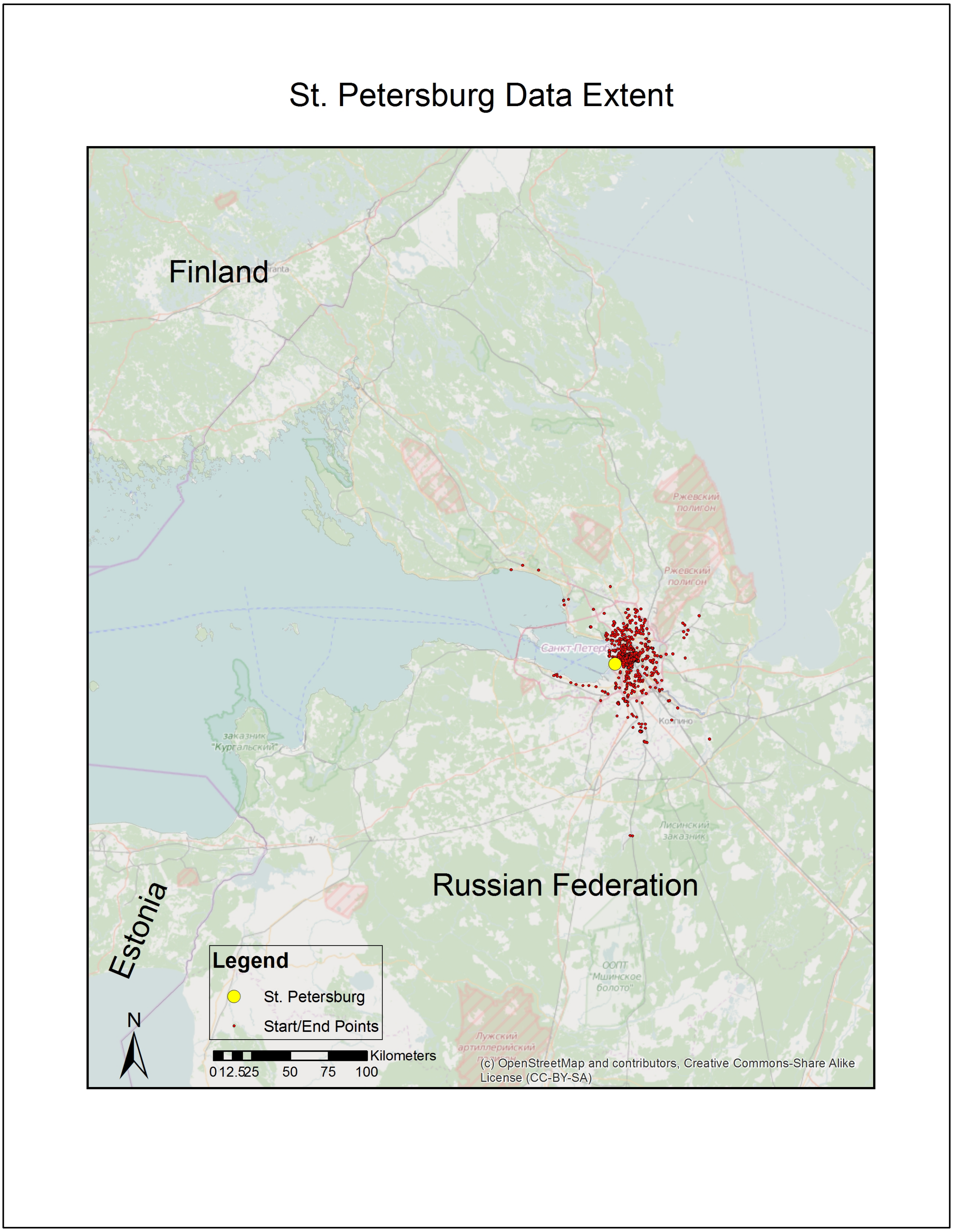

Figure 6. The full extent of the St. Petersburg data including both start and end points for both booked and searched rides.

3.3 City Profiles

3.3.1 London

London is the capital of the United Kingdom and is widely considered one of the top cities in the world by scholars and research institutes alike (Hales et al. 2014, Florida 2015, Derudder & Taylor 2012). In A.T. Kearney's 2014 report, London was ranked number two in the world based on factors such as business activity, human capital, information exchange, cultural experience and political engagement. According to the same report, London's absolute strength lies in its cultural experiences in the form of visual and performing arts, sporting events and international travelers (Hales et al. 2014).

Based on Florida's study, London was ranked the second most economically powerful city in the world (Florida 2015). Additionally, according to the GaWC (Global and World Cities) Research Network, London is one of *only* two cities in the entire world to receive an "Alpha ++" distinction (Derudder & Taylor 2012).

In 2013, the number of overseas visitors to London was highest since 1961, with a total of 16.8 million visits (Office for National Statistics 2014). In the beginning of 2015, the Greater London Authority reported a population of 8.6 million people with a population density of 5,491 people per square kilometer (Smith 2015).

3.3.2 Paris

Paris is the capital of France and is also widely considered as one of the top cities in the world. In A.T. Kearney's 2014 report, Paris was ranked number three in the world with its absolute strength coming from a high degree of information exchange particularly in the form of access to television news (Hales et al. 2014). Based on Florida's study, Paris was ranked as the fifth most economically powerful city in the world (Florida 2015). Also, in the GaWC Index, Paris was ranked one level below London and was considered an "Alpha +" city in 2012 (Derudder & Taylor 2012).

In 2013, the number of overseas visitors going to Paris was 15.5 million.

According to the yearly report published by the Paris Convention and Visitors Bureau, the current population of Paris is 2.2 million with a population density of 21,428 people per square kilometer (Paris Convention and Visitors Bureau 2013).

3.3.3 St. Petersburg

Contrary to the robust global positioning of both London and Paris, St. Petersburg does not even find itself in the 2014 A.T. Kearney report or other reputable global rankings (Florida 2015). St. Petersburg did however make its way into the GaWC Index, but only as third level city with “Gamma +” ranking (Derudder & Taylor 2012).

St. Petersburg has the second largest population of any Russian city with an estimated 4.8 million people in 2011 (Central Intelligence Agency 2011).

4. MATERIALS AND METHODS

4.1 Materials

4.1.1 Ride Data

All of the analyses performed in this thesis depend upon two core data sets, both of which were provided in-house by [redacted] Data Analyst, Juha Lehto.

The first data set includes all booked rides since the launch of [redacted] operations in [redacted] and continues up to the end of [redacted] when the analysis was performed. Automatic updates have been set up for the booked ride data and the file is updated once every 24 hours to the [redacted] internal server. A copy of the booked ride data was accessed from the [redacted] server at the end of [redacted].

The second data set includes all searched ride data since the launch of [redacted] operations in [redacted] and continues up to the end of [redacted] when the analysis was performed. Lehto especially prepared this latter data set for the purpose of this master's thesis according to the specifications requested. The searched ride data file is particularly large, so it was helpful that the data could be narrowed down by city using the supplier list captured in Table 2 above.

Unlike other studies (Chainey et al. 2008, Armstrong et al. 1999, Boulos et al. 2006), it was not deemed necessary to apply a geographic masking technique to preserve anonymity and confidentiality to either the booked or searched ride data. The coordinates used in the data are genuine and have not been displaced from the original location in any deliberate manner.

4.1.2 Booked Ride Data

The first data set needed to be massaged to correspond to the study area defined for this thesis. As the data set originally contained *all* booked rides for every [redacted] destination, data for every supplier except those operating in London, Paris and St. Petersburg needed to be excluded. Accordingly, just over 76% of the original [redacted] records were removed during this phase. Furthermore, an additional 1% of the booking data was removed. This 1% was made up of test bookings, bookings with missing coordinate data and [redacted] bookings that did not take place in London. ([redacted] provides service throughout the United Kingdom, so those rides taking place outside of London were excluded.)

The sum of booked rides for all three cities is [redacted] and individually there have been [redacted] booked rides in London, [redacted] booked rides in Paris and [redacted] booked rides in St. Petersburg.

4.1.3 Searched Ride Data

The second data set required a considerable amount of manual processing to get the data fit for analysis. The original data set consisted of [redacted] records, of which [redacted] applied to Paris, [redacted] applied to St. Petersburg and a whopping [redacted] applied to London.

Once split into the respective cities, the data needed to be combed further to eliminate unnecessary data. Some examples of data that needed to be removed include price page updates, offer comparison records, actual booking data, data from test booking channels, duplicate data and erroneous data. These categories are discussed in further detail below.

Price page updates are run automatically every week for the purpose of updating the “example rates” page of the [redacted] website [redacted] Offer comparison records are those generated in-house by users of the offer comparison tool. This tool allows users to find all the suppliers offering service from point A to point B and shows the net supplier rate for the fare.

All booked rides were also included in the searched ride data set. As the booked rides are already being considered in the first data set, they were removed. This was fortunately a straightforward operation, because the data included the genuine booking numbers for booked rides. Any piece of data with a genuine booking number was taken away. In the searched ride data, there was also data from test booking channels such as “apitest” and “developer”. These are channels used expressly by the [redacted] Development/Tech Team to evaluate and review new APIs or version releases, for example.

Duplicate data is an additional category of data that needed to be stripped away. For example, if you are looking for a ride from A to B in Paris and there are five unique offers, this shows up as five unique lines in the data. I am only interested in the search itself not all of the offers generated. If the duplicate records were not removed here, one route with ten offers would carry the same weight in the data as ten routes with one offer each. A majority of these duplicate data lines were removed by Lehto in the data preparation, however not all were removed. The remainder had to be manually extracted.

Finally, erroneous data also had to be removed. The biggest example from this category occurred in the London data with searches made through a specific channel, which repeatedly assigned the same erroneous coordinates (51.502922, 0.053314) for customer origin and destination points. If the channel could not find the location from the text field entered by the customer, it automatically set the coorindates to London City Airport (51.502922, 0.053314). This presence of this error was verified through discussions with the Development Team and through looking at the data to see a definite mismatch the textual input for the data and the coordinates selected.

All of the data categories above are either not genuine customer inquiries or are fundamentally flawed and if used would introduce error to the analysis. For these reasons, they have been removed from the final data set used in the analysis.

In the end, when all of the unnecessary data was stripped away, I was left with [redacted] total searched rides. Of these, there were [redacted] in London, [redacted] in Paris and [redacted] in St. Petersburg. Accordingly, just over 56% of the original London records, 26% of the original Paris records and 58% of the original St. Petersburg records were removed during the searched ride data processing phase.

Additionally, within the searched ride data, London was a bit of a special case. This is because of one supplier, [redacted], who provides service throughout the entire United Kingdom. As such, all records where the origin or destination did not have anything to do with London needed to be removed. To achieve this as efficiently as possible, I defined a coordinate grid around the London metropolitan area. Any piece of data whose start or end point lay within this grid was considered to be a part of London for the purpose of the ensuing data analysis. Any data whose start or end point that was not within this grid was removed. The grid was user defined and extends 52N, 51S, .50E and -.80W. Defining this grid help to structure the data processing and made it proceed more seamlessly.

4.1.4 Data Notes

Cleaning up data is a common practice in GIS so that the researcher is left with a solid core for analysis. In Grubesic's analysis of sex offender clusters in the state of Illinois, he removed nearly 60% of the records originally obtained from the Illinois State Police Sex Offender Information web site (Grubesic 2010).

While removing unnecessary data is essential, it is important to walk a fine line and not eliminate too much data. Scholars warn against this practice especially with respect to one's attribute data. They note that it is very easy to remove attribute data, however once the attribute data has been removed it is quite difficult to add again. The booked and searched ride data was processed with these above considerations in mind to guide the process and achieve an optimal result.

4.2 Methods

The data was split up into four distinct sets of data for each city, including;

a. Booking origin data
b. Booking destination data
c. Searched ride origin data
d. Searched ride destination data

Once the data was imported into ArcGIS, it was subjected to the Point Density tool within the Spatial Analyst toolbox. For the first data set analyzed for each city – *booked origin data* – the suggested settings for cell size and radius were used. In the output layer, the data was displayed with a one-third standard deviation. To enable equal comparison between all four visualizations of the same city, the same cell size, radius and class breaks were used. All of the output shapefiles have been added to a personal geodatabase created specifically for this project and the naming structure of the files was kept well organized for ease of use in the future. With respect to the map output, the standard considerations for work product were used, including legend, title, north arrow, sources, scale bar, labeling and more.

Processing the data for was the most intensive and time-consuming aspect of this thesis. The methods used were more straightforward and light when compared to the heaviness of the initial data processing.

5. RESULTS

Using the materials and methods outlined above, four point density analyses per city were created. This means that the finished library of work has yielded four maps each for London, Paris and St. Petersburg. For the ***booked*** ride data, two point density maps were created, one each for the origin and the destination data. This was replicated with the ***searched*** ride data with one point density map for both the origin and the destination data. All of the output maps can be found in Figures 7-18, located in this section.

5.1 Booked Ride Data

The point density of the booked ride data for all three cities was not surprising. The observed clustering pattern for both the origin and destination data matched my expectations. For example, in London, the highest point densities were observed around London Heathrow Airport, central London and London Gatwick Airport. To a lesser extent there was also activity around both London Stansted and Luton airports, which are major transportation hubs. The same was observed in Paris, with the highest densities observed at Charles du Gaulle Airport, Orly Airport and Paris city center. In the St. Petersburg case, the clustering was observed at the Finlyandsky Railway Station, Pulkovo Airport and St. Petersburg city center. These results match existing tabular statistics [redacted] has on these destinations.

5.2 Searched Ride Data

When looking at the searched ride data results, it is necessary to go back to the original hypothesis that was stated at the beginning of the thesis. I postulated there, *"that each city has hotspots that customers are interested in booking but current pricing inhibits these bookings from being made."* I was ideally looking for the data to point to specific routes in each city that need to be enhanced. The observations did not match the expected outcome, and each city actually behaved quite differently from one another. Although the expected outcome was not achieved, this ended up being an excellent result because the observation showed a diverse spectrum of city behavior within the [redacted] network.

Below, I discuss the results for each of the three cities in further detail. In this thesis, I have been systematically presenting the city order alphabetically (London, Paris, St.

Petersburg). However, due to the nature of the results, I will now present in order of complexity, starting with the least complex case (St. Petersburg), followed by an intermediate example (Paris) and ending with the most complex (London).

5.2.1 St. Petersburg

Of the three cities in this case study, the observed clustering in both the booked and searched ride data differed the least in St. Petersburg. In this city, the booked ride count actually exceeded the searched ride count – the only city within this case study to exhibit this particular characteristic. St. Petersburg has a very good conversion rate in the [redacted] network and offers are most often turning into actual bookings. From the clustering that was observed, there are not any significant observations of routes being searched but not booked. The booked ride data and the searched ride data match up quite well. In St. Petersburg, the pricing situation is quite healthy and in line with customer expectations for the market. The situation for St. Petersburg is ideal and it is in [redacted] best interest to have other cities performing this way to maximize volumes and profit.

5.2.2 Paris

The situation in Paris was quite different than what was observed in St. Petersburg. That being said, the observations in Paris were exactly in line with what I was originally expecting to see in each of these cities. There is a clearly defined mismatch between the booked ride data and the searched ride data and there are quite a few areas that are being searched but not booked. I discussed these results with stakeholders in the [redacted] Supply Team and I came up with a list of key areas that need to be analyzed and considered in further detail. The list has been included below in Table 3.

1	[redacted]
2	[redacted]
3	[redacted]
4	[redacted]
5	[redacted]
6	[redacted]

Table 3 – Top areas in Paris to be analyzed further based on the point density results.

5.2.3 London

The results of the London point density analysis were the most complex, but at the same time richest of the three cities in this case study. The differences between the booked ride data and the searched ride data were enormous and there are a staggering amount of searches being made that are not converting over to actual bookings. However, because there was so much data, it was much more difficult to hone in on small, specific areas as was possible in the Parisian case. In London I am left with the following observations/areas to be investigated in further detail. First, the central core of London is frequently booked, however there are significant searches around that central core that are not converting to bookings. Additionally, [redacted] and [redacted] are areas of concern along with [redacted], [redacted], [redacted], [redacted] and [redacted]. The data also gives a good indication of the overall London catchment area, extending as far north as [redacted] and less significantly, as far north as [redacted].

Although not as expected, the results of the three cities yielded extremely valuable data spanning a diverse spectrum of city performance within the [redacted] network.

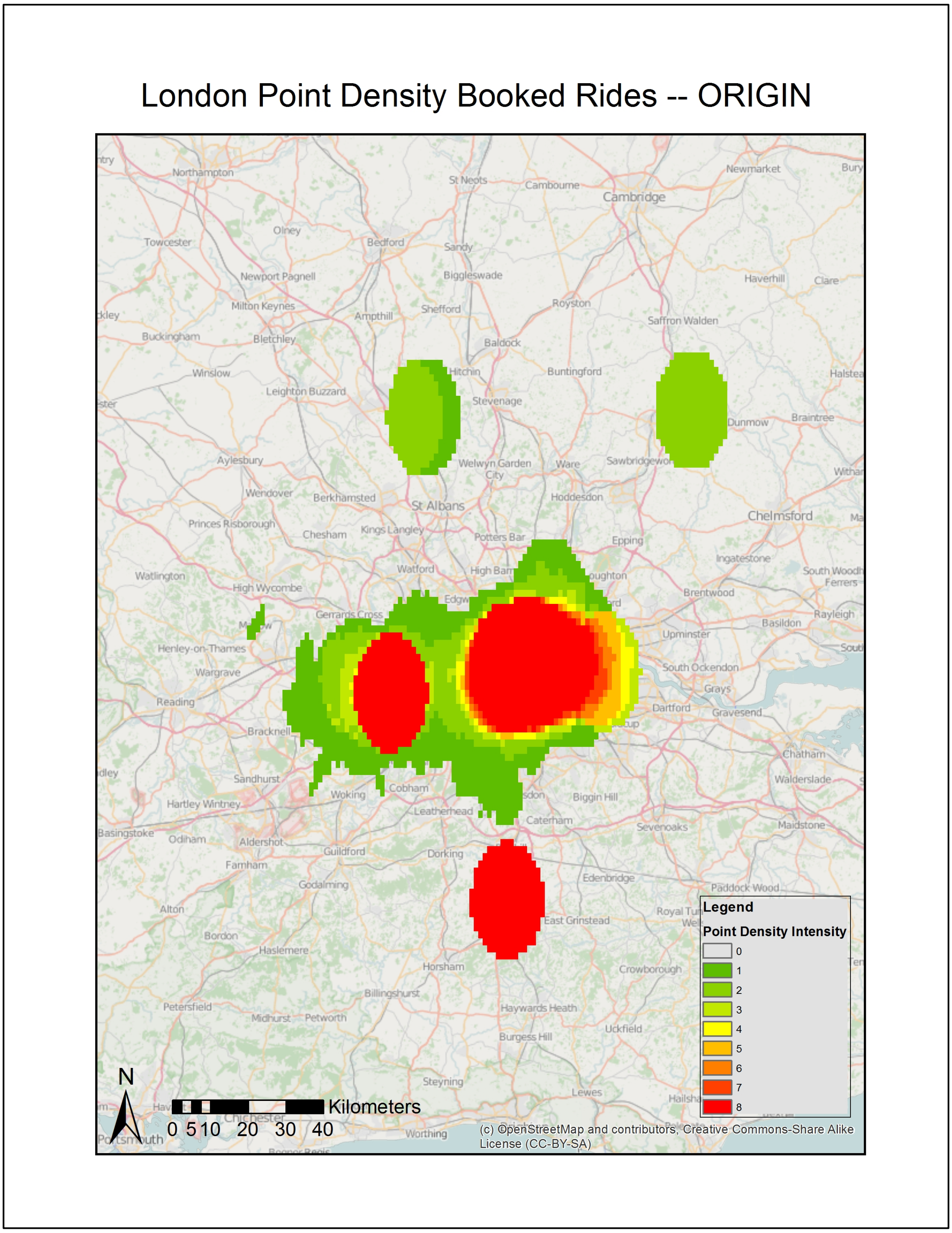

Figure 7. Point density map for all **booked** rides by **origin** in London from [redacted]

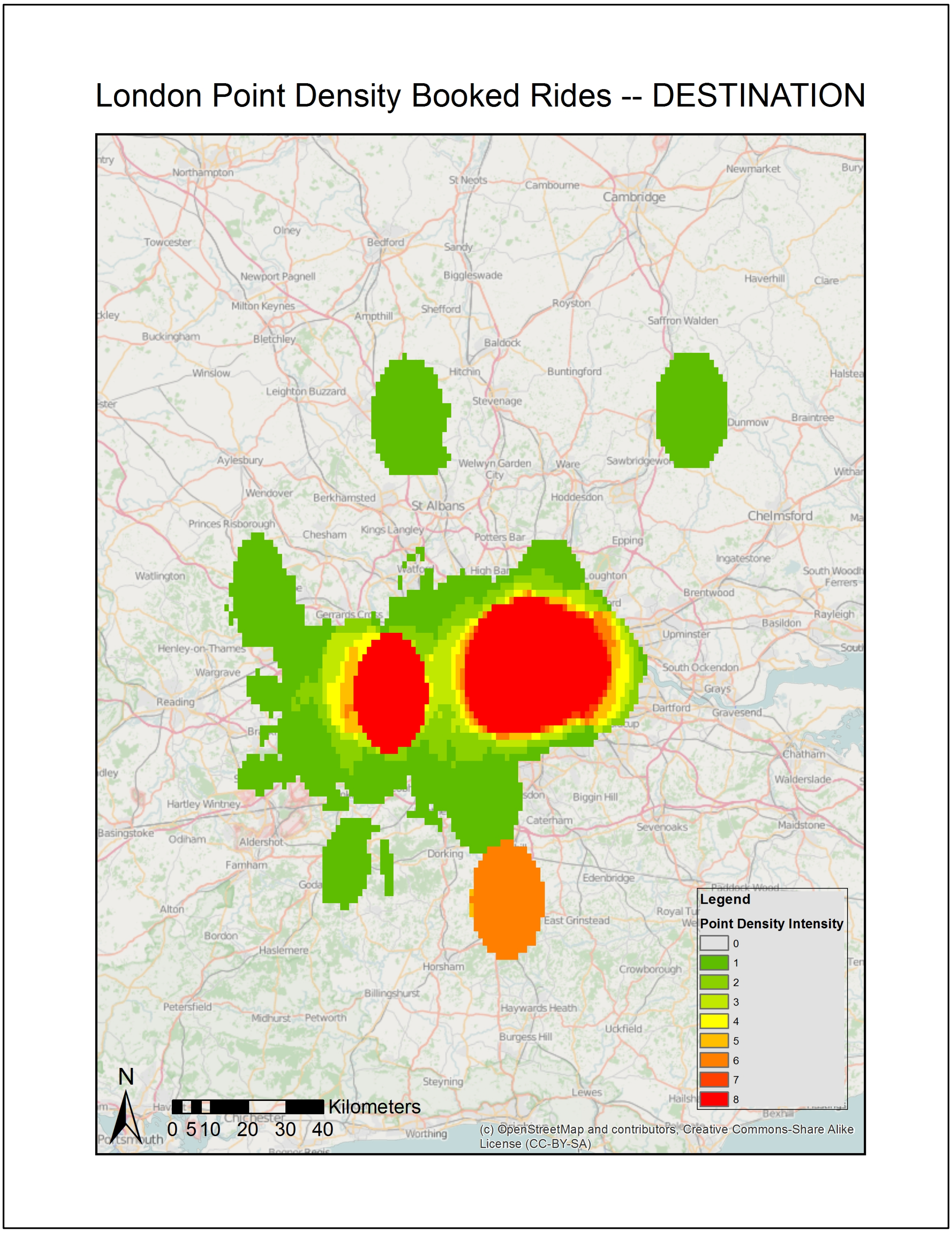

Figure 8. Point density map for all **booked** rides by **destination** in [redacted]

London Point Density Searched Rides -- ORIGIN

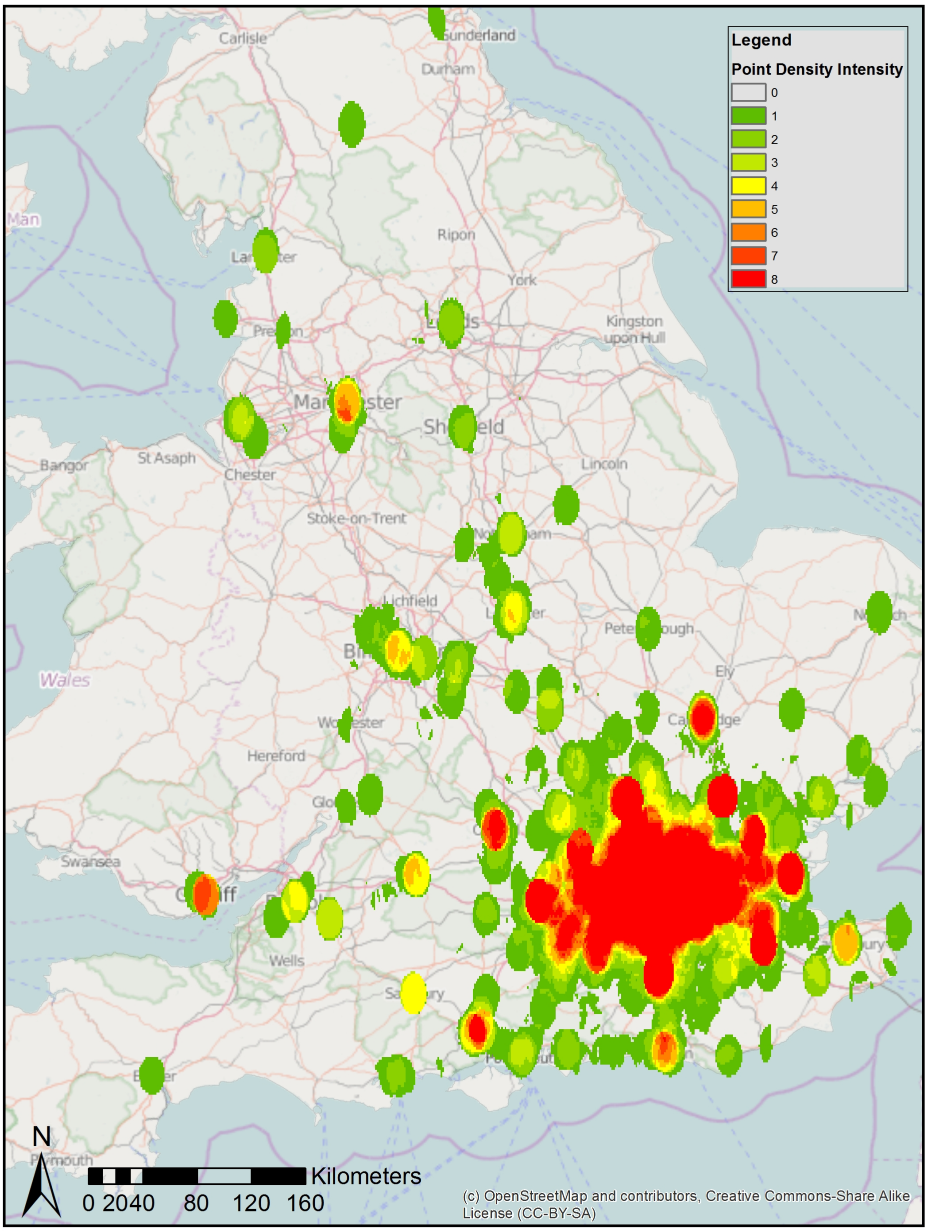

Figure 9. Point density map for all **searched** rides by **origin** in London [redacted]

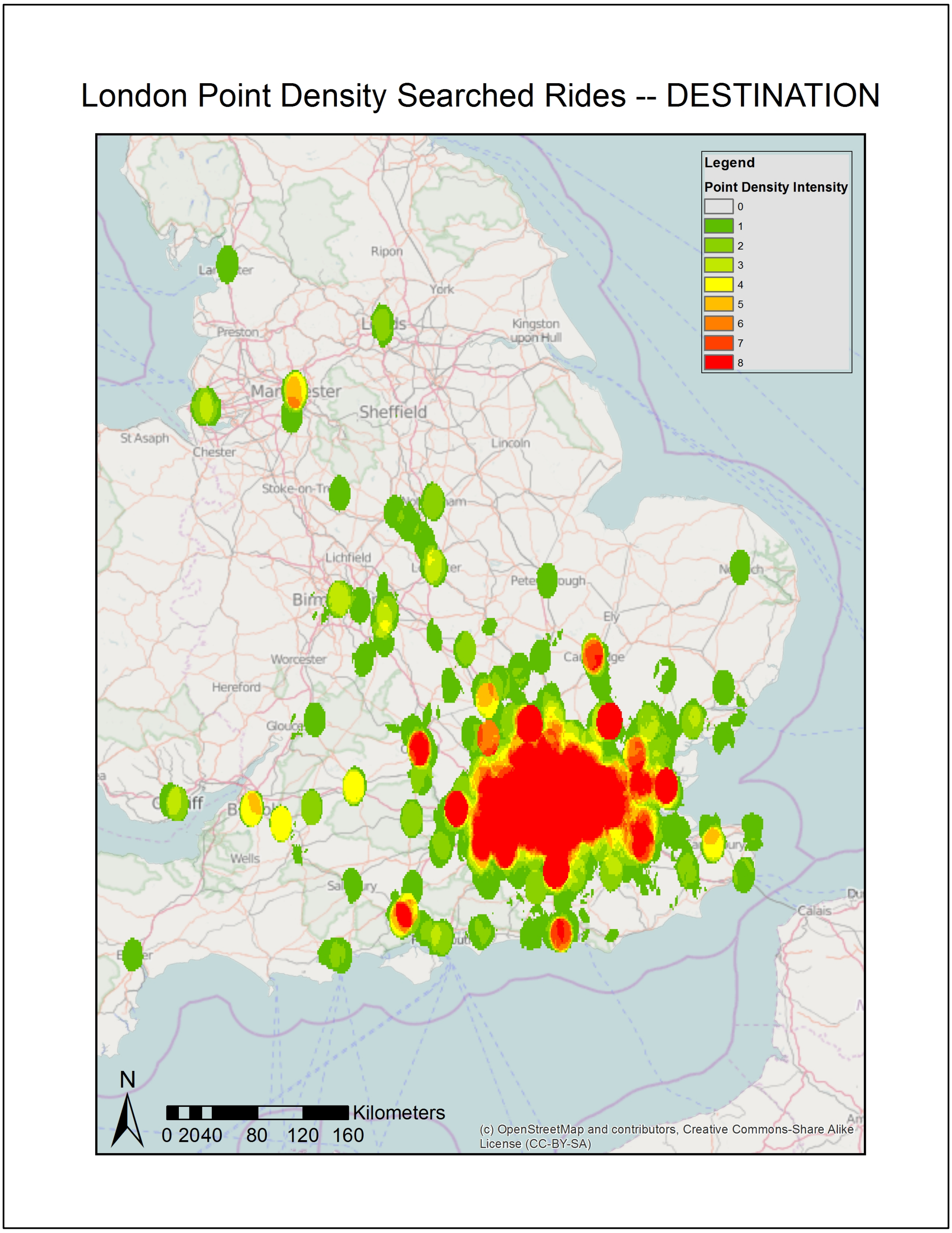

Figure 10. Point density map for all **searched** rides by **destination** in London [redacted]

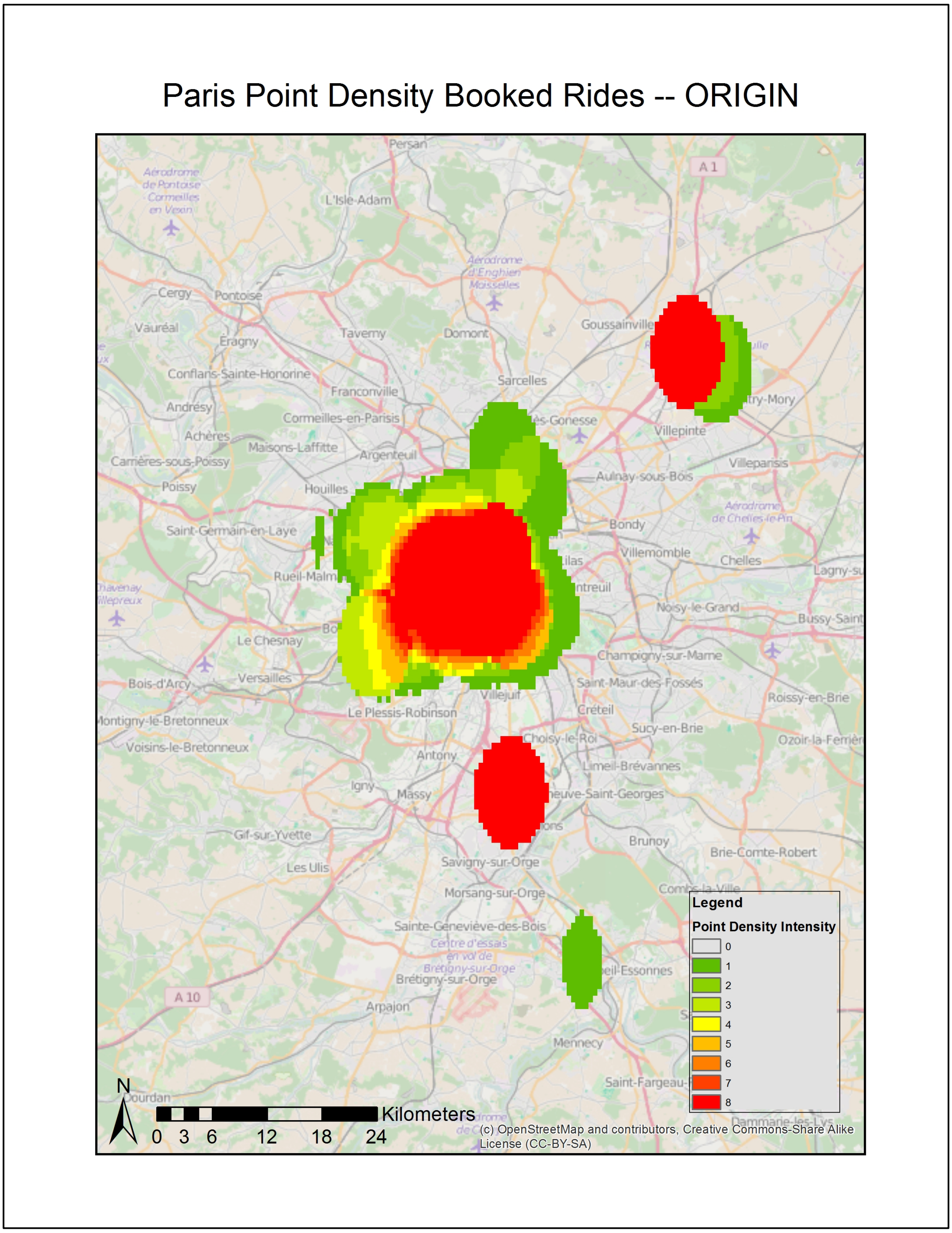

Figure 11. Point density map for all **booked** rides by **origin** in Paris from [redacted]

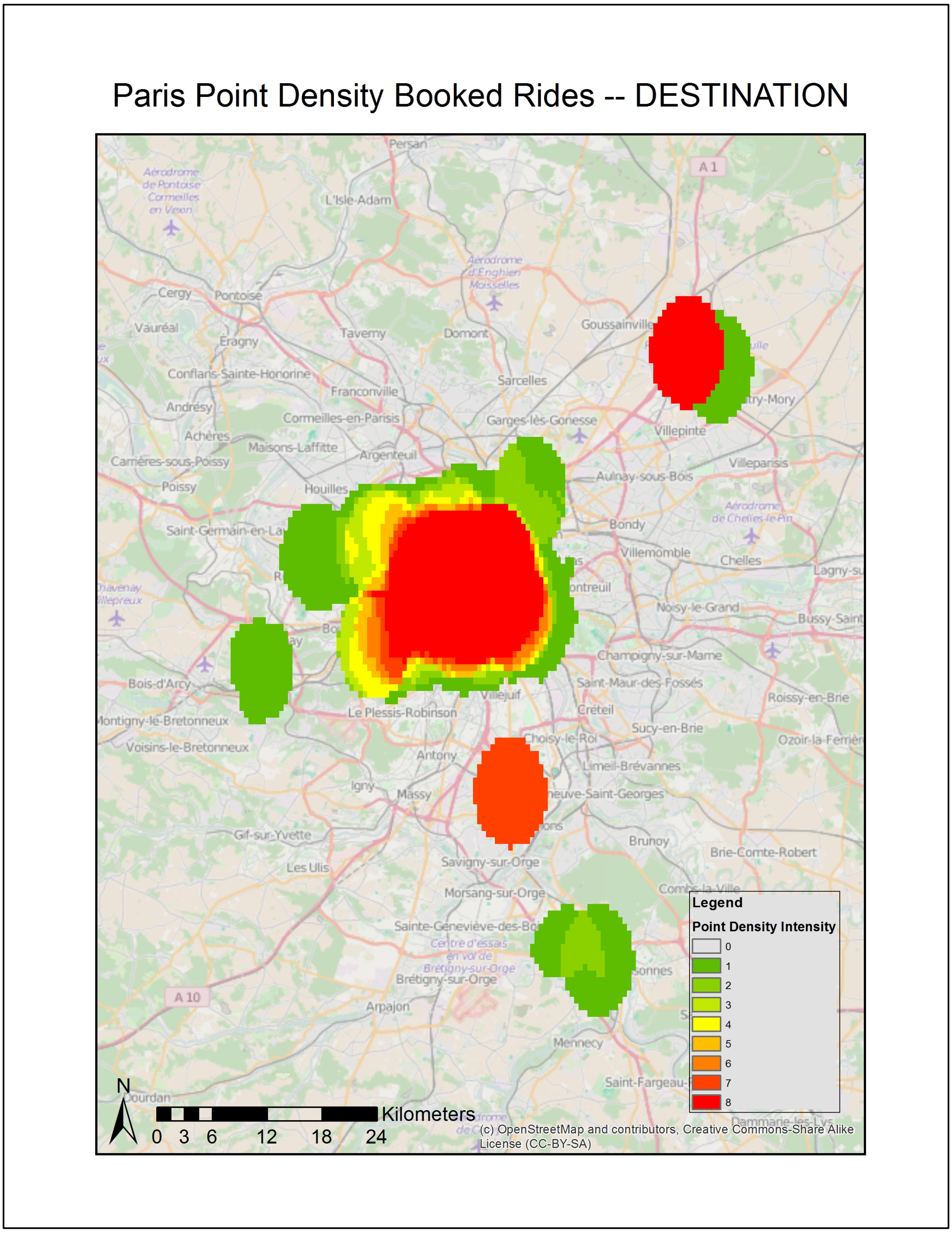

Figure 12. Point density map for all **booked** rides by **destination** in Paris from [redacted]

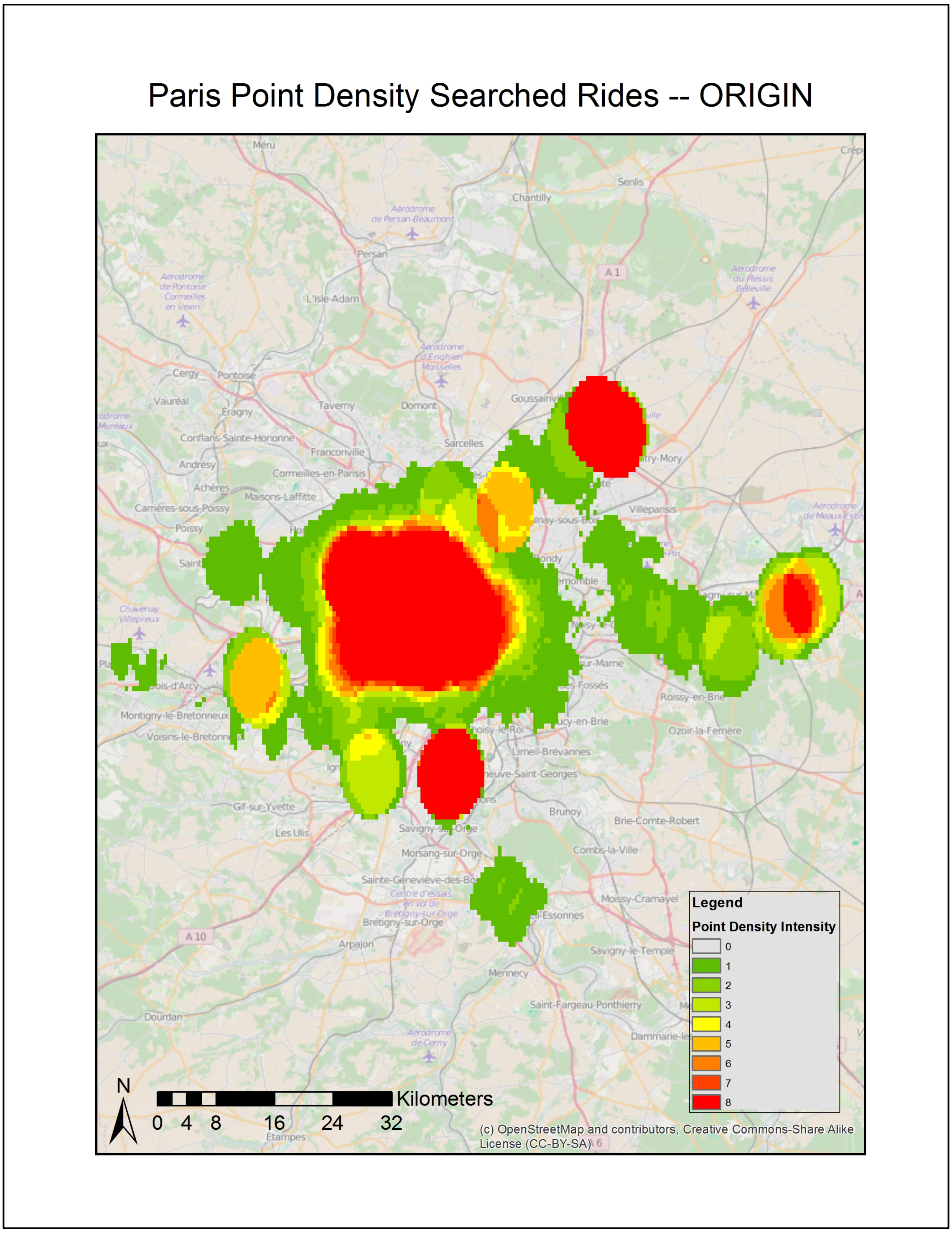

Figure 13. Point density map for all **searched** rides by **origin** in Paris from [redacted]

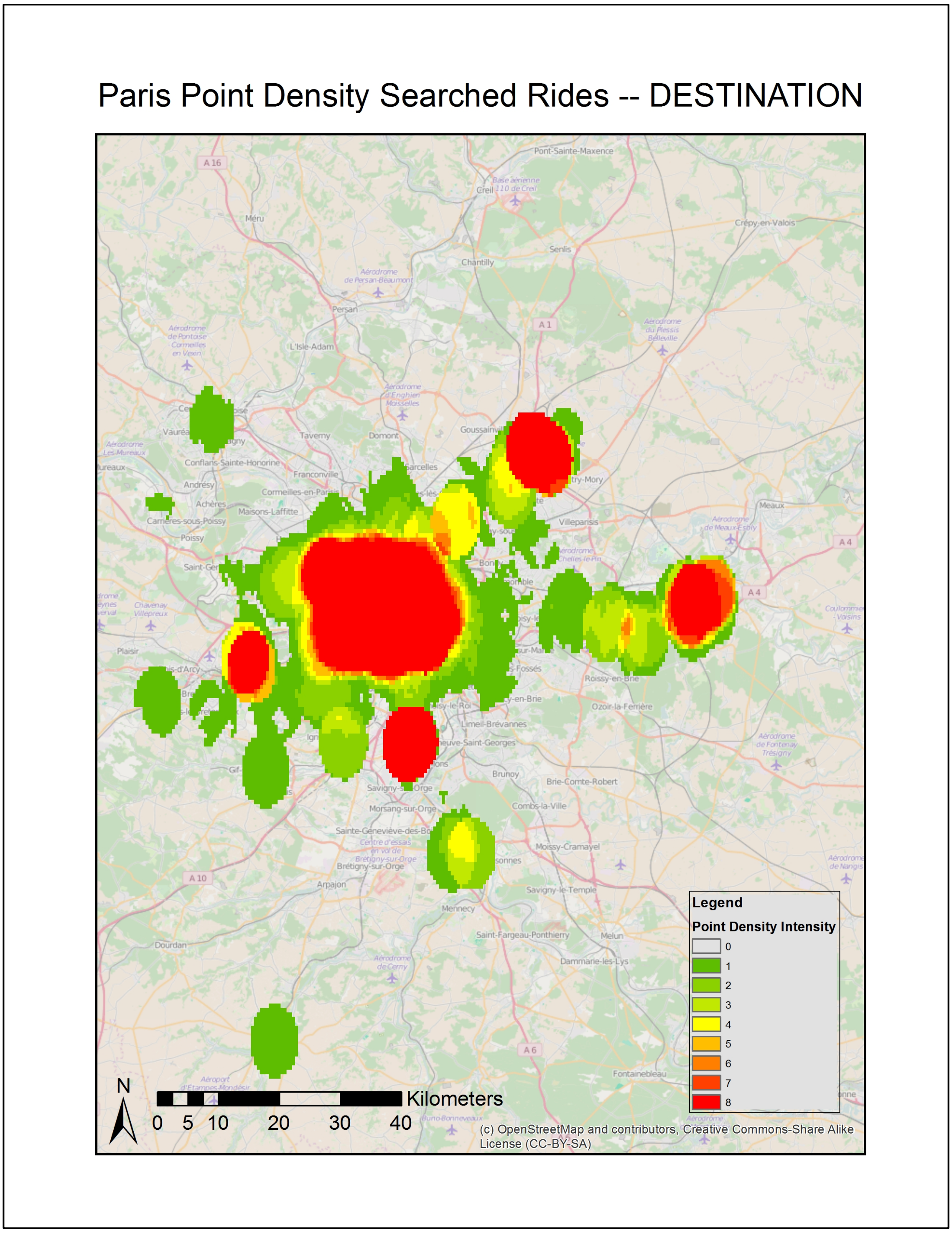

Figure 14. Point density map for all **searched** rides by **destination** in Paris from [redacted]

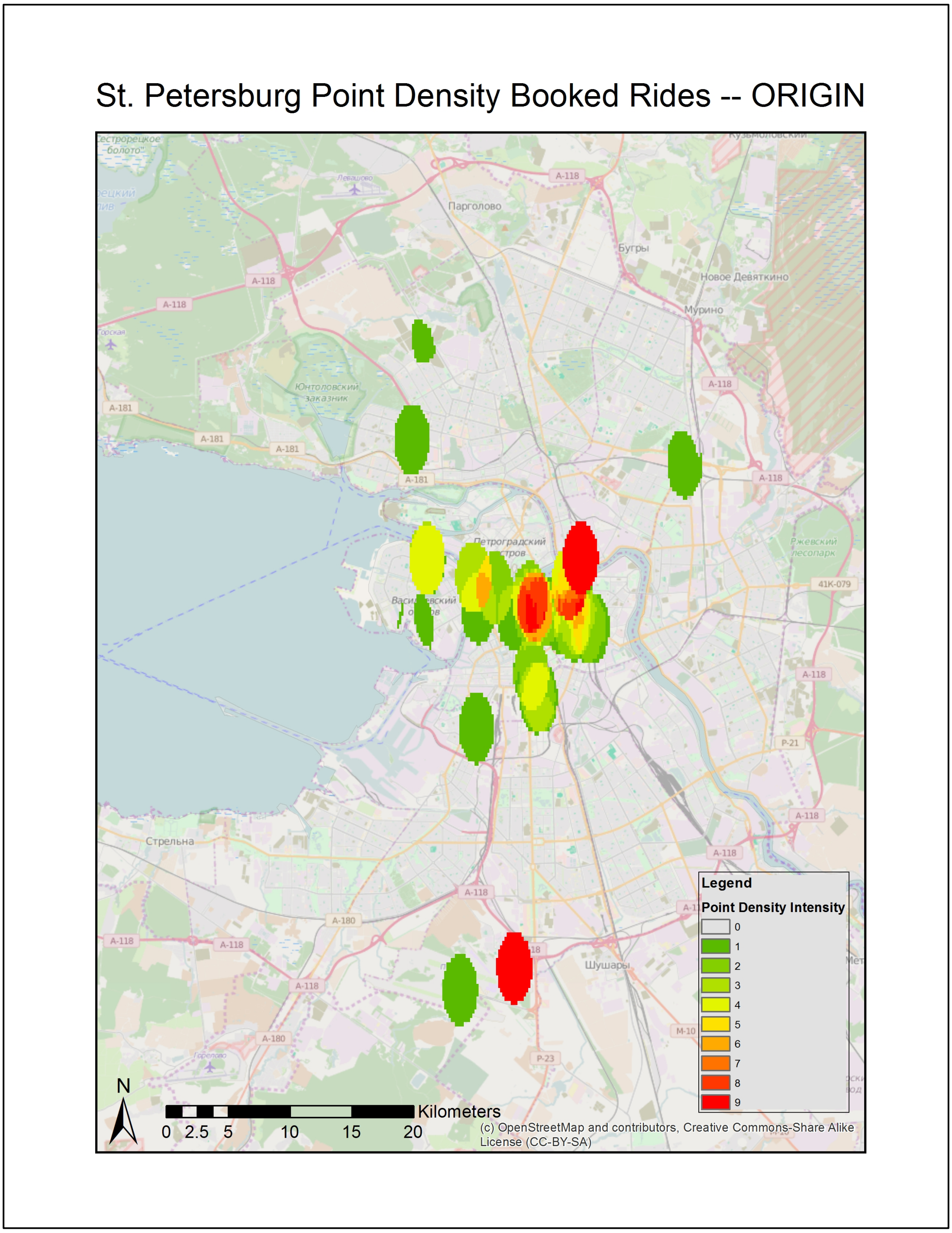

Figure 15. Point density map for all **booked** rides by **origin** in St. Petersburg from [redacted].

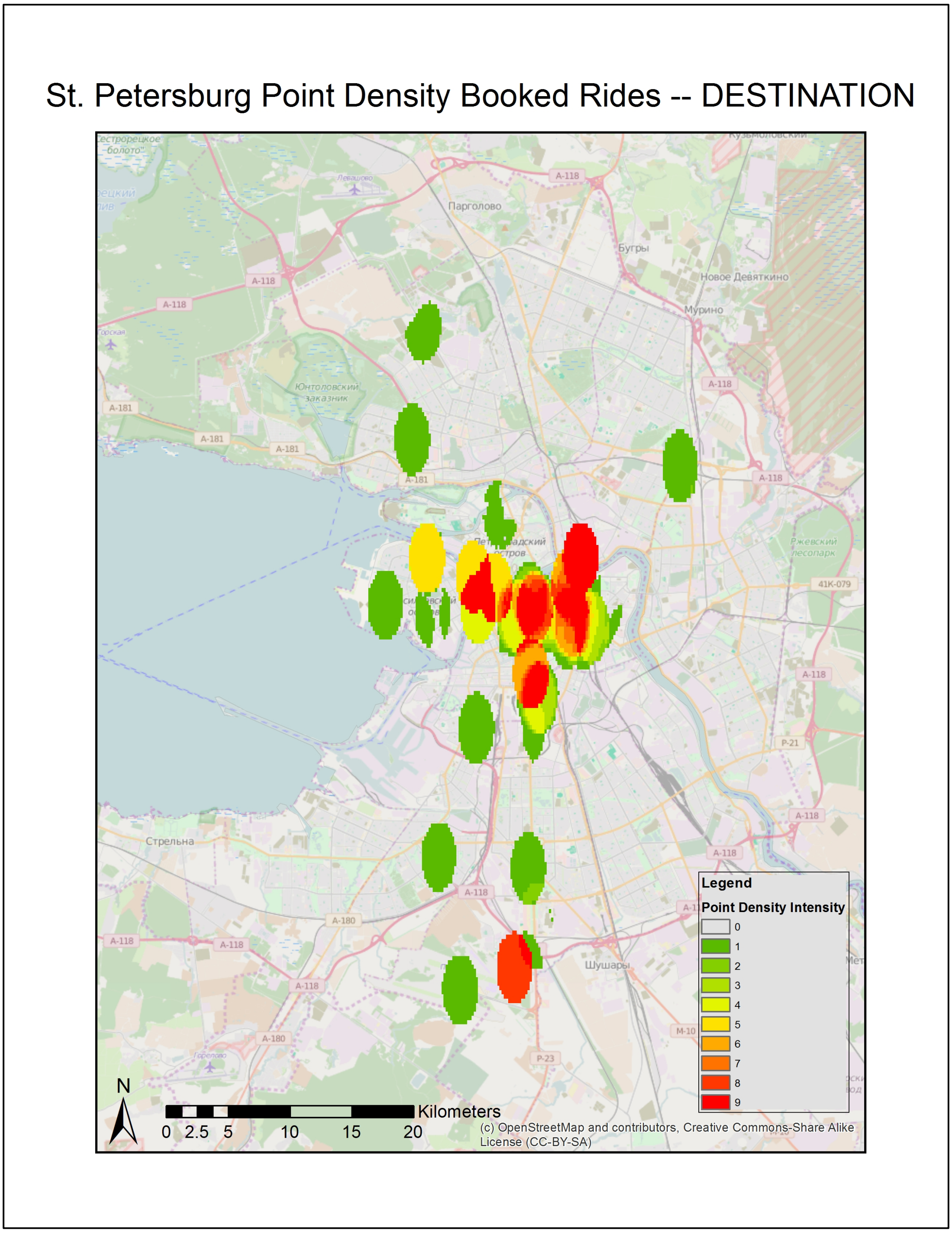

Figure 16. Point density map for all **booked** rides by **destination** in St. Petersburg from [redacted].

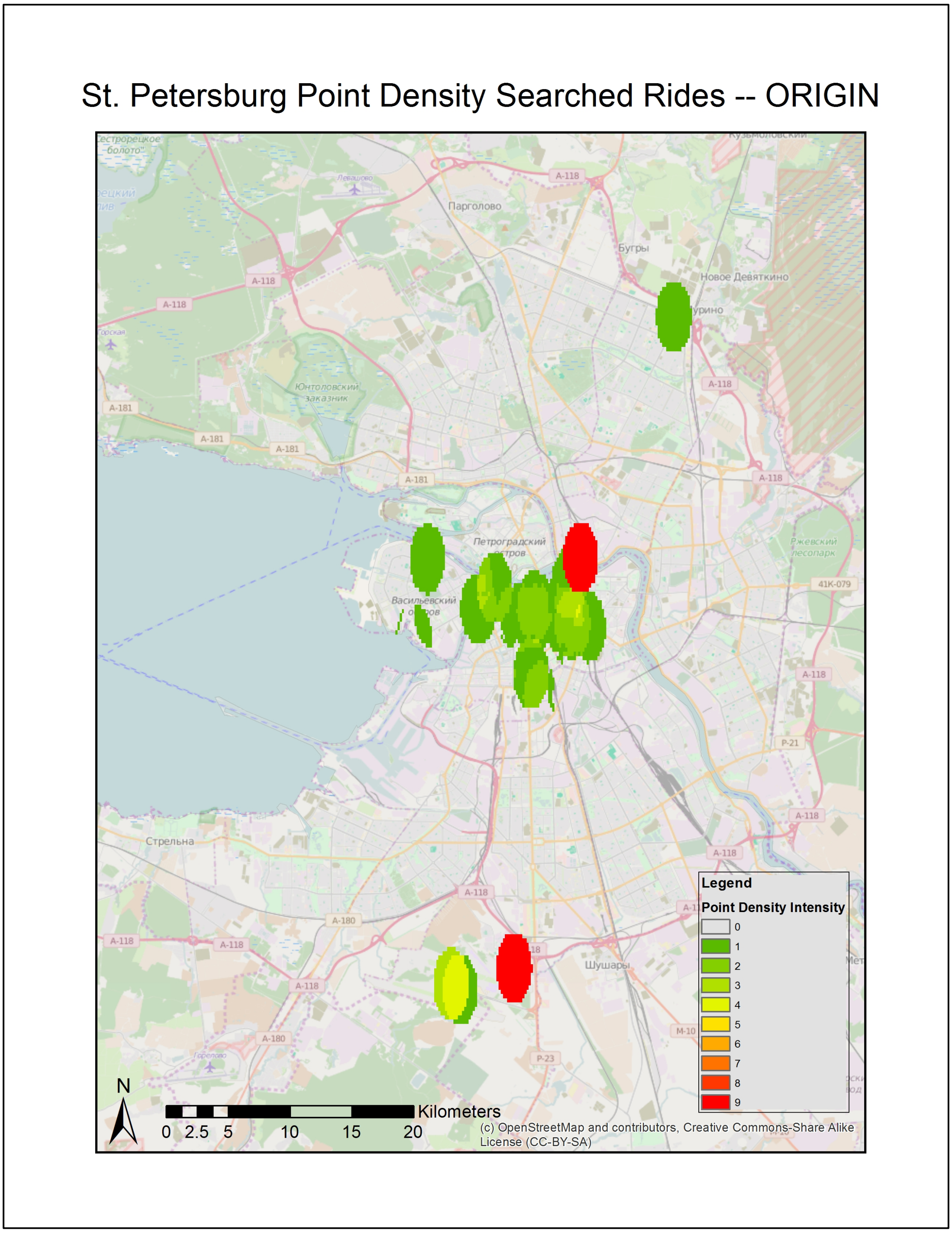

Figure 17. Point density map for all **searched** rides by **origin** in St. Petersburg from [redacted].

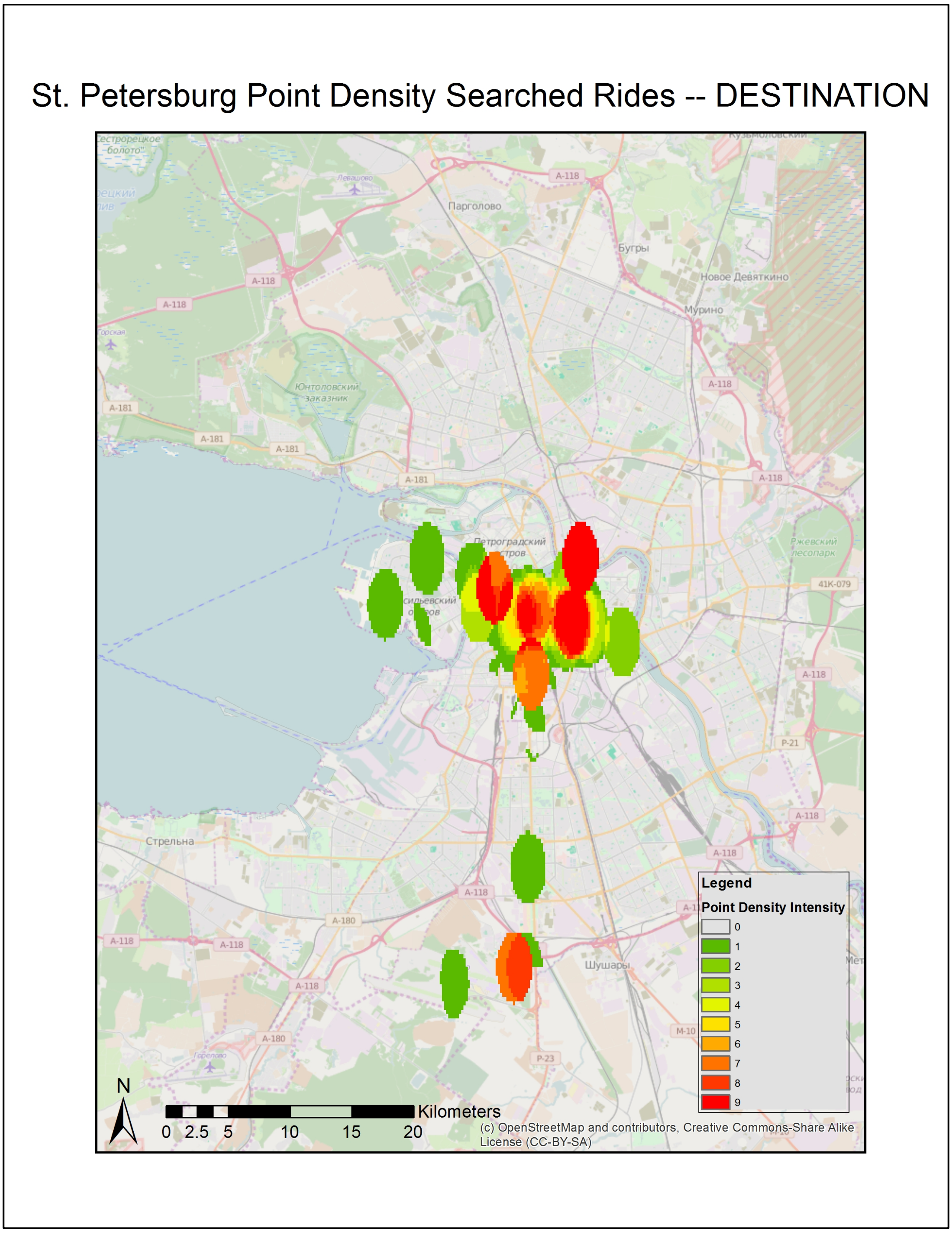

Figure 18. Point density map for all **searched** rides by **destination** in St. Petersburg from [redacted].

6. DISCUSSION

6.1 Outcome

As was stated, the results of this thesis did not match the initial expectations. I projected that I would be able to find specific routes/transfers to be improved in every city of the case study. In the end, Paris was the only city that behaved according to my expectations. I was initially upset with this outcome, however once I thought about and discussed the results in more detail, I came to realize how rich the offering actually was. The results of this analysis yield a model of city behavior within the [redacted] network.

Starting with St. Petersburg, there was no significant mismatch between the booked and searched ride data. This is not a bad thing, quite the contrary; St. Petersburg is behaving as a “model” city. To maximize volumes and profits, all cities need to be analyzed and unlocked, with the goal of having their data behave more like St. Petersburg, e.g. low degrees of mismatch between booked and searched ride data.

Within this model of city behavior, Paris performed right down the middle as an intermediate case. It is not yet a “model” city like St. Petersburg, nor is it extremely complex like the third actor in the model, London. As an intermediate city, there are clear and significant areas of mismatch between the booked and searched ride data. Some top areas in need of additional attention can be identified and these usually present themselves in the form of specific routes. Most enhancements should be able to be obtained through the network of existing suppliers.

Finally, London yielded quite a complex set of data and can be placed at the advanced end of the [redacted] network model. There are significant incidents of mismatch between the booked and searched ride data. However, the area is so big and the dataset is so large that it is not possible to pinpoint individual routes for enhancement. Instead, broader areas that need improvements can be identified. Some of these enhancements can be achieved with existing suppliers in the network; however, recruitment of additional supply will be necessary to make the city a better performer in terms of profit and volume.

The results did not match the hypothesis, but they did yield a fantastic model of city types within the [redacted] network: ideal, needs improvement and critical attention required.

6.2 Key Recommendations

A number of key recommendations can be given based on the results of this analysis. With respect to St. Petersburg, there is not anything specific that needs to take place, however, it is important to highlight that this is a "model" city. St. Petersburg is the benchmark for other cities and it should periodically be analyzed to ensure it is still having ideal performance.

In terms of key recommendations, Paris has quite a few specific routes to be unlocked and analyzed. As stated in the table in the results section, the key areas to be looked after in Paris include [redacted], [redacted], [redacted], [redacted], [redacted] and [redacted].

While not possible to make recommendations on specific areas/routes in London, it is possible to name broader areas that need to be addressed. These have also been previously listed. They include [redacted], [redacted], [redacted], [redacted], [redacted], [redacted], [redacted] and [redacted]. While it is certainly a viable option to reach out and engage with current suppliers, I strongly believe it will be necessary to get new suppliers on board who can fully support the demands and needs of these areas. Additionally, [redacted] and [redacted] just started a major ground transportation project in the UK, so these recommendations should be flagged to this team so they can approach any new suppliers matching these destination profiles. Another key recommendation with respect to London concerns the overall analysis of the data. In this case, it was quite heavy to interpret, so I would suggest breaking down the data for analysis into more sizable chunks for any of these "complex" cities going forward.

6.3 Data Concerns

One of the most time consuming and manual components of this thesis was stripping out London specific searched ride data from the nationwide [redacted] records. To enhance the situation going forward, a fundamental shift in the data reporting is called for. Right now, the data is presented at the supplier level with some indication of the

the data, it would have been extremely easy to only pull the data from the London based [redacted] products, for example, instead of having everything lumped together as they were in this exercise.

As mentioned previously in section **4.1.3.**, there were some difficulties working with the searched ride data as well as a result of the unique offers being made for each search. The duplicate data needed to be manually stripped away in these instances.

I have some small but lingering concerns about the overall legitimacy of the London searched ride data, particularly with respect to the erroneous data that had to be removed because of the London City Airport coordinate errors. First, I would like to make it clear that it was the best decision for this project was to remove the erroneous data given concerns about time management. If it had been left as is, the volumes to and from London City Airport would have looked much larger and more significant than they were. However, the proper way to make the data more robust and whole would have been by assigning the correct coordinates to the origin or the destination as needed. This would have been extremely time consuming to manually go through the 1000s of records, however it is one way to ensure a stronger set of data even if it is unknown how many of these records actually have something to do with London and would have been used in the final analysis in the end.

6.4 Future Work

Based on the results of this project, I believe that the case for spatial analysis within [redacted] has been adequately displayed and there is strong potential for future spatial analysis work. However, if [redacted] has the desire to replicate this analysis in other destinations or in the same destinations at a future date, some improvements with the data structure are called for, the main issues of which have been outlined above. Making these adjustments will significantly speed up the time necessary to perform this type of analysis.

Additionally, this type of work is time consuming and resource intensive, so any of the work needs to be carefully and strategically planned. Considering current resources, it would not make sense for a project of this scope to be performed on each and every destination. As such, future work in this area should run alongside the major goals of the company to yield the best results. For example, if it is quarter one and the main trade and airline focus will be on Germany during quarter two, it would be well within the company strategy to orchestrate an analysis on the main German cities to guide supply and pricing decisions that will support the upcoming company goals. Projects of this scope do require careful planning and a certain idea of where the future demands will be concentrated.

Other items to consider with respect to future work include the period of time and the city profile model. In the analysis done here, the time frame runs from [redacted] through [redacted]. In the future, should the entire data set again be considered or only a few months? Was there a major incident and if so, should the analysis look at the situation before and after this incident? As the analysis is potentially rolled out to other cities, does the city profile model suggested here remain intact or does it need to be adjusted to include a fourth, fifth or sixth profile type?

In the end, even though the thesis itself felt quite heavy and cumbersome, the resulting spatial analysis was quite rewarding. I am looking forward to driving the key recommendations further within the organization to enact change and positively affect the [redacted] growth.

ACKNOWLEDGEMENTS

I would be lying if I said working on this thesis was a pleasure. It was not. I am absolutely delighted by the fact that this is now over.

I couldn't have gotten past the finish line without the help and support of some key individuals.

Darryn Quirk, thanks for being my main study buddy and for the frequent meet-ups and generous encouragement. Thanks also to Oxana Kozar for her feedback and guidance while we slogged through this together.

A very big heartfelt thank you to both Leena Malkki of the Network for European Studies and Markku Löytönen of the Department of Geography for all of your patience, encouragement and again patience.

Thanks also to everyone at [redacted], especially Andreas Hansson and Petteri Niemi, who provided me with a phenomenal opportunity to link my studies with my work. Thanks also to Juha Lehto for giving me all the data I needed and Mika Friman for answering my questions with respect to data structure and content.

To my parents Thomas and Anna Riley, thank you for showing me the importance of an education at an early age and for providing me with so many opportunities to succeed. I am blessed to have the both of you in my life.

And finally, thank you to my Matti for all of your encouragement and for nudging me forward even when I did not want to be nudged and expressly told you to leave me alone. You knew what was best for me despite my many protestations!

REFERENCES

Alden, W. (2014). The Business Tycoons of Airbnb. The New York Times. 11.03.15 < http://www.nytimes.com/2014/11/30/magazine/the-business-tycoons-of-airbnb.html?_r=0>

Armstrong, M., G. Rushton & D. Zimmerman (1999). Geographically Masking Health Data to Preserve Confidentiality. *Statistics in Medicine* 18: 5, 497-525.

Balogun, T., H. Okeke & C. Chukwukere (2014). Crime Mapping in Nigeria Using GIS. *Journal of Geographic Information System* 2014:6, 453-466.

Blackburn, J., T. Hadfield, A. Curtis & M. Hugh-Jones (2014). Spatial and Temporal Patterns of Anthrax in White-Tailed Deer, Odocoileus virginianus, and Hematophagous Flies in West Texas during the Summertime Anthrax Risk Period. *Annals of the Association of American Geographers* 104: 5, 939-958.

Blaschke, T. & H. Merschdorf (2014). Geographic information science as a multidisciplinary and multiparadigmatic field. *Geography and Geographic Information Science* 41: 3, 196-213.

Boulos, M., Q. Cai, J. Padget & G. Rushton (2006). Using software agents to preserve individual health data confidentiality in micro-scale geographic analyses. *Journal of Biomedical Informatics.* 39: 2, 160-170.

[redacted] (2012). [redacted] SIGNED. 8pp. Helsinki, Finland.

[redacted] (2012). [redacted] SIGNED. 8pp. Helsinki, Finland.

[redacted] (2012). [redacted] Feb 2012 SIGNED. 15pp. Helsinki, Finland.

[redacted] (2013). [redacted] 17.06.2013 SIGNED. 11pp. Helsinki, Finland.

[redacted] (2013). [redacted] April 17 2013 SIGNED. 14pp. Helsinki, Finland.

[redacted] (2013). [redacted] 8_2013 SIGNED. 10pp. Helsinki, Finland.

[redacted] (2014). [redacted] 4_2014 SIGNED. 15pp. Helsinki, Finland.

[redacted] (2014). [redacted] 3_2014 SIGNED. 11pp. Helsinki, Finland.

[redacted] (2014). [redacted] 11_2014_SIGNED. 12pp. Helsinki, Finland.

[redacted] (2014). [redacted] 4_2014 SIGNED. 12pp. Helsinki, Finland.

[redacted] (2014). [redacted] 11_2014_SIGNED. 12pp. Helsinki, Finland.

[redacted] (2015). [redacted] 2_2015 SIGNED. 13pp. Helsinki, Finland.

[redacted] (2015). [redacted] 1_2015 SIGNED. 11pp. Helsinki, Finland.

Cambie, G., N. Sanchez-Carnero, T. Mingozzi, R. Muino & J. Freire (2013). Identifying and mapping local bycatch hotspots of loggerhead sea turtles using a GIS-based method: implications for conservation. *Marine Biology* 160: 2013, 653-665.

Central Intelligence Agency (2011). The World Factbook 2014. 12.03.2015 < https://www.cia.gov/library/publications/the-world-factbook/geos/rs.html>

Chainey, S. (2012). *Understanding Hotspots*. Workshop presented at Australian Crime Mapping and Analysis Conference, Melbourne, Australia. https://www.ucl.ac.uk/scs/people/academic-research-staff/spencer-chainey/Slides/Melbourne2012_Understanding_hotspots

Chainey, S., L.Tompson & S. Uhlig (2008). The Utility of Hotspot Mapping for Predicting Spatial Patterns of Crime. *Security Journal* 21: 2008, 4-28.

Cheng, E., H. Li & L. Yu (2007). A GIS approach to shopping mall location selection. *Building and Environment.* 42: 2007, 884-892.

Cord, D. (2015). [redacted] Attracts Major Acquirer with [redacted]. Good News from Finland. 03.11.15 <http://www.goodnewsfinland.com/feature/[redacted]-attracts-major-acquirer-with-[redacted]/>

Curtis, A., X. Ye, E. Heob, J. Targhetta, V. Salvato, M. Reyna, R. Bueno Jr. & L. Holmes (2014). A comparison of three approaches to idenfity West Nile Virus mosquito space-time hotspots in the Houston Vicinity for the period 2002-2011. *Applied Geography* 51: 2014, 58-64.

Cutler, M. (2013). [redacted] Goes After the Corporate Taxi Booking Market with 30% Growth Month-Over-Month. Tech Crunch. 09.03.15. < http://techcrunch.com/2013/06/04/[redacted]-2/>

Derudder, B., & P. Taylor (2012). The World According to GaWC 2012. Globalization and World Cities Research Network. 25.03.15 < http://www.lboro.ac.uk/gawc/world2012t.html>

Eck, J., S. Chainey, J. Cameron, M. Leitner & R. Wilson (2005). *Mapping Crime: Understanding Hot Spots* 79 pp. U.S. Department of Justice – Office of Justice Programs, Washington, D.C.

Eckert, J.& S.Shetty (2011). Food systems, planning and quantifying access: Using GIS to plan for food retail. *Applied Geography* 31: 2011, 1216-1223.

Eckley, D.C. & K.M. Curtin (2012). Evaluating the spatiotemporal clustering of traffic incidents. *Computers, Environment and Urban Systems* 37 :2013, 70-81.

Florida, R. (2015). Sorry, London: New York is the World's Most Economically Powerful City. The Atlantic - City Lab. 26.03.15. < http://www.citylab.com/work/2015/03/sorry-london-new-york-is-the-worlds-most-economically-powerful-city/386315/>

Fox, L. (2013). [redacted] Looks to Drive Down Cost of Taxis Across Europe Via Pre-Booked Model. Tnooz. 09.03.15 < http://www.tnooz.com/article/[redacted] looks-to-drive-down-cost-of-taxis-across-europe-via-pre-booked-model/>

Grubesic, T. (2010). Sex offender clusters. *Applied Geography* 30: 2010, 2-18.

Grubesic, T., R. Wei & A. Murray (2014). Spatial Clustering Overview and Comparison: Accuracy, Sensitivity and Computational Expense. *Annals of the Association of American Geographers* 104: 2014, 1134-1155.

Hales, M., A. Peña, E. Peterson & J. Gott (2014). *2014 Global Cities Index and Emerging Cities Outlook.* 16pp. A.T. Kearney, Chicago.

Hall, T., P. Toms, M. McGuinness, C. Parker & N. Roberts (2015). Where's the Geography Department? The changing administrative place of Geography in UK higher education. *Area* (Royal Geographic Society with the Institute of British Geographers) 47: 1, 56-64.

Hansson, A. (2014). We call it Cabregation. [redacted] 27.03.2015 <http://www.[redacted]-travel.com/2014/08/cabregation/>

Harries, K. (1999). *Mapping Crime Principles and Practice.* 193pp. US Department of Justice – Office of Justice Programs, Washington D.C.

Holmgren, T. (2014). Destination Driven: [redacted] Lands at South Africa's 'Big 5' Airports. [redacted].com 27.03.2015 <http://www.[redacted]-travel.com/2014/09/destination-driven-[redacted]-lands-south-africas-big-5-airports/>

Kasimov, N., S. Chalov & A. Panin (2013). Multidisciplinary field training in undergraduate Physical Geography: Russian experience. *Journal of Geography in Higher Education* 37: 3, 416-431.

Kiseleva, V. (2015). Now on the Map – Bulgaria, Morocco & Taiwan, also in – more destinations & price cuts. www.[redacted]-travel.com. 02.11.15 <https://www.[redacted]-travel.com/2015/10/now-on-the-map-bulgaria-morocco-taiwan-also-in-more-destinations-and-price-cuts/>

Kiseleva, V. (2015). Service Expansion in Greece, price drops in Russia and more destination updates. www.[redacted]-travel.com. 02.11.15 https://www.[redacted]-travel.com/2015/07/service-expansion-in-greece-price-drops-in-russia-and-more-destination-updates/

Macmillan, D., M. Spector & E. Rusli (2014). Airbnb Weighs Employee Stock Sale at $13 Billion Valuation. The Wall Street Journal. 11.03.15 <http://www.wsj.com/articles/airbnb-mulls-employee-stock-sale-at-13-billion-valuation-1414100930>

Mishra, S. (2009). GIS in Indian Retail Industry – A Strategic Tool. *International Journal of Marketing Studies* 1:1, 50-57.

Montreuil, B. (2011). Toward a Physical Internet: Meeting the Global Logistics Sustainability Grand Challenge. *Logistics Research* 3:2, 71-87.

Niemi, P. (2015). [redacted] Booking Data. Received 17.02.2015.

Office for National Statistics (2013). Travel Trends, 2013. 26.03.2015 < http://www.ons.gov.uk/ons/rel/ott/travel-trends/2013/sty.html>

Paris Convention and Visitors Bureau (2013). Tourism in Paris – Key Figures 2013. 18.03.2015 < http://asp.zone-secure.net/v2/index.jsp?id=1203/1515/45590&lng=en>

Raeste, J. (2015). [redacted] Myytiin Irlantiin. Helsingin Sanomat. 03.11.15 < http://www.hs.fi/talous/a1430801755249>

Ramadan, A., C. Lochhead & D. Peterson (2014). Behind Uber's Soaring Value. Fortune. 11.03.15 <http://fortune.com/2014/12/11/behind-ubers-soaring-value/>

Riley, J. (2015). [redacted] Expands Throughout Asia. [redacted]. 09.03.15 <http://www.[redacted]-travel.com/2015/02/[redacted]-expands-throughout-asia/>

Riley, J. (2015). [redacted] Destinations Update. [redacted] 09.03.15 <http://www.[redacted]-travel.com/2015/02/[redacted]-destinations-update/>

Riley, J, jennifer.riley[redacted], 2015. Chiang Mai, Chiang Rai, Ko Samui & Krabi, Thailand Now Live!. [E-mail] Message to CE [redacted] Employees (kaikki@[redacted].com). Sent October 15th, 15:09. Available at: jennifer.riley@[redacted]com [Accessed 02 November 15]

Silberman, J., & P. Rees (2010). Reinventing mountain settlements: A GIS model for identifying possible ski towns in the U.S. Rocky Mountains. *Applied Geography* 30: 2010, 36-49.

Smith, D. (2015). London's Population High: Top Metropolis Facts. BBC News. 20.03.2015 < http://www.bbc.com/news/uk-england-london-31056626>

Tobin, K., L. Hester, M. Davey-Rothwell & C. Latkin (2012). An Examination of Spatial Concentrations of Sex Exchange Norms among Drug Users in Baltimore, Maryland. *Annals of the Association of American Geographers* 105: 2012, 1058-1066.

van Haaren, R., & V. Fthenakis (2011). GIS-based wind farm site selection using spatial multi-criteria analysis (SMCA): Evaluating the case for New York State. *Renewable and Sustainable Energy Reviews* 15: 2011, 3332-3340.

Vuori, M. (2009). *Accessibility as a Determinant of Opportunities: A Case Study from Peruvian Amazonia.* Master's Thesis. 87pp. Department of Geography University of Helsinki.

Wing, M., & J. Tynon (2006). Crime Mapping and Spatial Analysis in National Forests. *Journal of Forestry* 104: 2006, 293-298.